



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07H 21/02, 21/04, C12N 15/00, C12P 21/00, C07K 14/00, A61K 35/14, 38/00, C12Q 1/00		A1	(11) International Publication Number: WO 98/23631 (43) International Publication Date: 4 June 1998 (04.06.98)
(21) International Application Number: PCT/US97/21976			19401 (US). ZARFOS, Philip, N. [US/US]; 1907 Yorktown North, Norristown, PA 19403 (US).
(22) International Filing Date: 24 November 1997 (24.11.97)			(74) Agents: GIMMI, Edward, R. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).
(30) Priority Data: 60/031,879 27 November 1996 (27.11.96) US			(81) Designated States: CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
(71) Applicants (for all designated States except US): SMITHKLINE BEECHAM CORPORATION [US/US]; One Franklin Plaza, Philadelphia, PA 19103 (US). SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).			Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(72) Inventors; and			
(75) Inventors/Applicants (for US only): BLACK, Michael, Terence [GB/US]; 502 Milhouse Way, Chester Springs, PA 19425 (US). HODGSON, John, Edward [GB/US]; 260 Lapp Road, Malvern, PA 19355 (US). KNOWLES, David, Justin, Charles [GB/GB]; Ladywell House, York Road, Boroughbridge, North Yorkshire YO5 9EB (GB). LONETTO, Michael, Arthur [GB/GB]; 18 Victoria Circle, Collegeville, PA 19426 (US). NICHOLAS, Richard, Oakley [GB/US]; 355 Carmen Drive, Collegeville, PA 19426 (US). REID, Robert, H., Jr. [US/US]; 8 Pacer Lane, Norristown, PA			

(54) Title: NOVEL BACTERIAL POLYPEPTIDES AND POLYNUCLEOTIDES

(57) Abstract

This invention relates to newly identified Streptococcal polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

NOVEL BACTERIAL POLYPEPTIDES AND POLYNUCLEOTIDES**FIELD OF THE INVENTION**

This invention relates to newly identified polynucleotides and polypeptides, and their production and uses, as well as their variants, agonists and antagonists, and their uses. In particular, in these and in other regards, the invention relates to novel polynucleotides and polypeptides set forth in Table 1.

BACKGROUND OF THE INVENTION

The Streptococci make up a medically important genera of microbes known to cause several types of disease in humans, including otitis media, pneumonia and meningitis. Since its isolation more than 100 years ago, *Streptococcus pneumoniae* (herein *S. pneumoniae*) has been one of the more intensively studied microbes. For example, much of our early understanding that DNA is, in fact, the genetic material was predicated on the work of Griffith and of Avery, Macleod and McCarty using this microbe. Despite the vast amount of research with *S. pneumoniae*, many questions concerning the virulence of this microbe remain.

While certain Streptococcal factors associated with pathogenicity have been identified, e.g., capsule polysaccharides, peptidoglycans, pneumolysins, PspA Complement factor H binding component, autolysin, neuraminidase, peptide permeases, hydrogen peroxide, IgA1 protease, the list is certainly not complete. Further very little is known concerning the temporal expression of such genes during infection and disease progression in a mammalian host. Discovering the sets of genes the bacterium is likely to be expressing at the different stages of infection, particularly when an infection is established, provides critical information for the screening and characterization of novel antibacterials which can interrupt pathogenesis. In addition to providing a fuller understanding of known proteins, such an approach will identify previously unrecognised targets.

GUG is used as an initiating nucleotide, rather than ATG, for a significant number of mRNA's in both Gram positive and Gram negative bacteria. Statistics on the frequency of NTG codons in the start codon for several bacterial species are available on line via computer at http://biochem.otago.ac.nz:800/Transterm/home_page.html.

A discussion of initiation codons in *B. subtilis* is set forth in Vellanoweth, RL.1993 in *Bacillus subtilis and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics*, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711. Vellenworth indicates a major difference between *B. subtilis* and the gram-negative organisms is in the choice of initiation codon. 91% of the sequenced *E. coli*

genes start with AUG. By contrast, about 30% of *B. subtilis* and other clostridial branch genes start with UUG or GUG. Moreover, CUG functions as a start codon in *B. subtilis*. Mutations of an AUG initiation codon to GUG or UUG often cause decreased expression in *B. subtilis* and *E. coli*. Generally, translation efficiency is higher with AUG initiation codons. A strong Shine-Delgarno ribosome binding site, however, can compensate almost fully for a weak initiation codon. It has been reported that genes with a range of expression levels have initiation codons other than ATG in gram positives (Vellanoweth, RL.1993 in Bacillus subtilis and other Gram Positive Bacteria, Biochemistry, Physiology and Molecular Genetics, Sonenshein, Hoch, Losick Eds. Amer. Soc. Microbiol, Washington DC. p. 699-711).

Provided herein are ORF sequences from genes possessing GUG initiation codons and proteins expressed therefrom and homologues thereto to be used for screening for antimicrobial compounds. Clearly, there is a need for polypeptide and polynucleotide sequences that may be used to screen for antimicrobial compound and which may also be used to determine the roles of such sequences in pathogenesis of infection, dysfunction and disease. There is also need, therefore, for identification and characterization of such sequences which may play a role in preventing, ameliorating or correcting infections, dysfunctions or diseases.

The polypeptides of the invention have amino acid sequence homology to a known protein(s) as set forth in Table 1.

SUMMARY OF THE INVENTION

It is an object of the invention to provide polypeptides that have been identified as novel polypeptides by homology between an amino acid sequence selected from the group consisting of the sequences set out in Table 1 and a known amino acid sequence or sequences of other proteins such as the protein identities listed in Table 1.

It is a further object of the invention to provide polynucleotides that encode novel polypeptides, particularly polynucleotides that encode polypeptides of *Streptococcus pneumoniae*.

In a particularly preferred embodiment of the invention the polynucleotide comprises a region encoding a polypeptide comprising a sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In another particularly preferred embodiment of the invention there is a novel protein from *Streptococcus pneumoniae* comprising an amino acid sequence selected from the group consisting of the sequences set out in Table 1, or a variant of any of these sequences.

In accordance with another aspect of the invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Streptococcus pneumoniae* 0100993 strain contained in the deposited strain.

A further aspect of the invention there are provided isolated nucleic acid molecules encoding a polypeptide of the invention, particularly *Streptococcus pneumoniae* polypeptide, and including mRNAs, cDNAs, genomic DNAs. Further embodiments of the invention include biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

In accordance with another aspect of the invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization. Among the particularly preferred embodiments of the invention are naturally occurring allelic variants of a polypeptide of the invention and polypeptides encoded thereby.

Another aspect of the invention there are provided novel polypeptides of *Streptococcus pneumoniae* as well as biologically, diagnostically, prophylactically, clinically or therapeutically useful variants thereof, and compositions comprising the same.

Among the particularly preferred embodiments of the invention are variants of the polypeptides of the invention encoded by naturally occurring alleles of their genes.

In a preferred embodiment of the invention there are provided methods for producing the aforementioned polypeptides.

In accordance with yet another aspect of the invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents, including, for example, antibodies.

In accordance with certain preferred embodiments of the invention, there are provided products, compositions and methods for assessing expression of the polypeptides and polynucleotides of the invention, treating disease, for example, including, for example, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema and endocarditis, and most particularly meningitis, such as for example infection of cerebrospinal fluid, assaying genetic variation, and administering a polypeptide or polynucleotide of the invention to an organism to raise an immunological response against a bacteria, especially a *Streptococcus pneumoniae* bacteria.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided polynucleotides that hybridize to a polynucleotide sequence of the invention, particularly under stringent conditions.

In certain preferred embodiments of the invention there are provided antibodies against polypeptides of the invention.

In other embodiments of the invention there are provided methods for identifying compounds which bind to or otherwise interact with and inhibit or activate an activity of a polypeptide or polynucleotide of the invention comprising: contacting a polypeptide or polynucleotide of the invention with a compound to be screened under conditions to permit binding to or other interaction between the compound and the polypeptide or polynucleotide to assess the binding to or other interaction with the compound, such binding or interaction being associated with a second component capable of providing a detectable signal in response to the binding or interaction of the polypeptide or polynucleotide with the compound; and determining whether the compound binds to or otherwise interacts with and activates or inhibits an activity of the polypeptide or polynucleotide by detecting the presence or absence of a signal generated from the binding or interaction of the compound with the polypeptide or polynucleotide.

In accordance with yet another aspect of the invention, there are provided agonists and antagonists of the polypeptides and polynucleotides of the invention, preferably bacteriostatic or bacteriocidal agonists and antagonists.

In a further aspect of the invention there are provided compositions comprising a polynucleotide or a polypeptide of the invention for administration to a cell or to a multicellular organism.

Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following descriptions and from reading the other parts of the present disclosure.

GLOSSARY

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Disease(s)" means any bacterial infection, but preferably a streptococcal infection, such as, otitis media, conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural empyema, endocarditis, meningitis, and infection of cerebrospinal fluid.

"Host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings

of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Preferred methods to determine identity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, the GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., *J. Molec. Biol.* 215: 403-410 (1990). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., *J. Mol. Biol.* 215: 403-410 (1990). As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "identity" to a reference nucleotide sequence it is intended that the nucleotide sequence of the tested polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. These mutations of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence. Analogously, by a polypeptide having an amino acid sequence having at least, for example, 95% identity to a reference amino acid sequence is intended that the test amino acid sequence of the polypeptide is identical to the reference sequence except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the reference amino acid. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino

acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

"Isolated" means altered "by the hand of man" from its natural state, *i.e.*, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living organism is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein.

"Polynucleotide(s)" generally refers to any polyribonucleotide or polydeoxyribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotide(s)" include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded regions, or a mixture of single- and double-stranded regions. In addition, "polynucleotide" as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term "polynucleotide(s)" also includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotide(s)" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term "polynucleotide(s)" as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including, for

example, simple and complex cells. "Polynucleotide(s)" also embraces short polynucleotides often referred to as oligonucleotide(s).

"Polypeptide(s)" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds. "Polypeptide(s)" refers to both short chains, commonly referred to as peptides, oligopeptides and oligomers and to longer chains generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene encoded amino acids. "Polypeptide(s)" include those modified either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. Modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, selenylation, sulfation, transfer-RNA mediated addition of amino acids to proteins, such as arginylation, and ubiquitination. See, for instance, *PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993) and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626-646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Polypeptides may be branched or cyclic, with or without branching. Cyclic, branched and branched circular polypeptides may result from post-translational natural processes and may be made by entirely synthetic methods, as well.

“Variant(s)” as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques, by direct synthesis, and by other recombinant methods known to skilled artisans.

DESCRIPTION OF THE INVENTION

Each of polynucleotide and polypeptide sequences provided herein may be used in the discovery and development of antibacterial compounds. Upon expression of the sequences with the appropriate initiation and termination codons the encoded polypeptide can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding preferably the amino terminal regions of the encoded protein or the Shine-Delgarno region can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The

first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. Because each of the sequences contains an open reading frame (ORF) with an appropriate initiation and termination codons, the encoded protein upon expression can be used as a target for the screening of antimicrobial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

It is believed that bacteria possess a number of ways of regulating gene expression levels, especially in subtle degrees, and the interplay between ribosome binding site and initiation codon is utilized for this purpose for these genes. It is also believed that such genes will be important targets for antimicrobial drug discovery, particularly since pathogenesis genes are believed undergo gene expression regulation during in the pathogenesis process. Therefore, the invention provides ORF sequences possessing a GTG (GUG) initiation codon and protein targets expressed therefrom.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches in related organisms. Thus, if a homologue exists, the Streptococcal-like form of this gene would likely play an analogous role. For example, a Streptococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

ORF Gene Expression

Recently techniques have become available to evaluate temporal gene expression in bacteria, particularly as it applies to viability under laboratory and infection conditions. A number of methods can be used to identify genes which are essential to survival *per se*, or essential to the establishment/maintenance of an infection. Identification of an ORF unknown by one of these methods yields additional information about its function and permits the selection of such an ORF for further development as a screening target. Briefly, these approaches include:

1) Signature Tagged Mutagenesis (STM): This technique is described by Hensel *et al.*, Science 269: 400-403(1995), the contents of which is incorporated by reference for background purposes. Signature tagged mutagenesis identifies genes necessary for the establishment/maintenance of infection in a given infection model.

The basis of the technique is the random mutagenesis of target organism by various means (e.g., transposons) such that unique DNA sequence tags are inserted in close proximity to the site of mutation. The tags from a mixed population of bacterial mutants and bacteria recovered from an infected hosts are detected by amplification, radiolabeling and hybridisation analysis. Mutants attenuated in virulence are revealed by absence of the tag from the pool of bacteria recovered from infected hosts.

In *Streptococcus pneumoniae*, because the transposon system is less well developed, a more efficient way of creating the tagged mutants is to use the insertion-duplication mutagenesis technique as described by Morrison *et al.*, J. Bacteriol. 159:870 (1984) the contents of which is incorporated by reference for background purposes.

2) In Vivo Expression Technology (IVET): This technique is described by Camilli *et al.*, Proc. Nat'l. Acad. Sci. USA. 91:2634-2638 (1994), the contents of which is incorporated by reference for background purposes. IVET identifies genes up-regulated during infection when compared to laboratory cultivation, implying an important role in infection. ORF identified by this technique are implied to have a significant role in infection establishment/maintenance.

In this technique random chromosomal fragments of target organism are cloned upstream of a promoter-less recombinase gene in a plasmid vector. This construct is introduced into the target organism which carries an antibiotic resistance gene flanked by resolvase sites. Growth in the presence of the antibiotic removes from the population those fragments cloned into the plasmid vector capable of supporting transcription of the recombinase gene and therefore have caused loss of antibiotic resistance. The resistant pool

is introduced into a host and at various times after infection bacteria may be recovered and assessed for the presence of antibiotic resistance. The chromosomal fragment carried by each antibiotic sensitive bacterium should carry a promoter or portion of a gene normally upregulated during infection. Sequencing upstream of the recombinase gene allows identification of the up regulated gene.

3) Differential display: This technique is described by Chuang *et al.*, J. Bacteriol. 175:2026-2036 (1993), the contents of which is incorporated by reference for background purposes. This method identifies those genes which are expressed in an organism by identifying mRNA present using randomly-primed RT-PCR. By comparing pre-infection and post infection profiles, genes up and down regulated during infection can be identified and the RT-PCR product sequenced and matched to ORF 'unknowns'.

4) Generation of conditional lethal mutants by transposon mutagenesis:

This technique, described by de Lorenzo, V. *et al.*, Gene 123:17-24 (1993); Neuwald, A. F. *et al.*, Gene 125: 69-73(1993); and Takiff, H. E. *et al.*, J. Bacteriol. 174:1544-1553(1992), the contents of which is incorporated by reference for background purposes, identifies genes whose expression are essential for cell viability.

In this technique transposons carrying controllable promoters, which provide transcription outward from the transposon in one or both directions, are generated. Random insertion of these transposons into target organisms and subsequent isolation of insertion mutants in the presence of inducer of promoter activity ensures that insertions which separate promoter from coding region of a gene whose expression is essential for cell viability will be recovered. Subsequent replica plating in the absence of inducer identifies such insertions, since they fail to survive. Sequencing of the flanking regions of the transposon allows identification of site of insertion and identification of the gene disrupted. Close monitoring of the changes in cellular processes/morphology during growth in the absence of inducer yields information on likely function of the gene. Such monitoring could include flow cytometry (cell division, lysis, redox potential, DNA replication), incorporation of radiochemically labeled precursors into DNA, RNA, protein, lipid, peptidoglycan, monitoring reporter enzyme gene fusions which respond to known cellular stresses.

5) Generation of conditional lethal mutants by chemical mutagenesis: This technique is described by Beckwith, J., Methods in Enzymology 204: 3-18(1991), the contents of which are incorporated herein by reference for background purposes. In this technique random chemical mutagenesis of target organism, growth at

temperature other than physiological temperature (permissive temperature) and subsequent replica plating and growth at different temperature (e.g. 42°C to identify ts, 25°C to identify cs) are used to identify those isolates which now fail to grow (conditional mutants). As above close monitoring of the changes upon growth at the non-permissive temperature yields information on the function of the mutated gene. Complementation of conditional lethal mutation by library from target organism and sequencing of complementing gene allows matching with unknown ORF.

6) RT-PCR: *Streptococcus pneumoniae* messenger RNA is isolated from bacterial infected tissue e.g. 48 hour murine lung infections, and the amount of each mRNA species assessed by reverse transcription of the RNA sample primed with random hexanucleotides followed by PCR with gene specific primer pairs. The determination of the presence and amount of a particular mRNA species by quantification of the resultant PCR product provides information on the bacterial genes which are transcribed in the infected tissue. Analysis of gene transcription can be carried out at different times of infection to gain a detailed knowledge of gene regulation in bacterial pathogenesis allowing for a clearer understanding of which gene products represent targets for screens for novel antibacterials. Because of the gene specific nature of the PCR primers employed it should be understood that the bacterial mRNA preparation need not be free of mammalian RNA. This allows the investigator to carry out a simple and quick RNA preparation from infected tissue to obtain bacterial mRNA species which are very short lived in the bacterium (in the order of 2 minute halflives). Optimally the bacterial mRNA is prepared from infected murine lung tissue by mechanical disruption in the presence of TRIzole (GIBCO-BRL) for very short periods of time, subsequent processing according to the manufacturers of TRIzole reagent and DNAase treatment to remove contaminating DNA. Preferably the process is optimised by finding those conditions which give a maximum amount of *Streptococcus pneumoniae* 16S ribosomal RNA as detected by probing Northerns with a suitably labelled sequence specific oligonucleotide probe. Typically a 5' dye labelled primer is used in each PCR primer pair in a PCR reaction which is terminated optimally between 8 and 25 cycles. The PCR products are separated on 6% polyacrylamide gels with detection and quantification using GeneScanner (manufactured by ABI).

Each of these techniques may have advantages or disadvantage depending on the particular application. The skilled artisan would choose the approach that is the most relevant with the particular end use in mind.

Use of the of these technologies when applied to the ORFs of the present invention enables identification of bacterial proteins expressed during infection, inhibitors of which would have utility in anti-bacterial therapy.

The invention relates to novel polypeptides and polynucleotides as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of *Streptococcus pneumoniae*, which is related by amino acid sequence homology to known polypeptide as set forth in Table 1. The invention relates especially to compounds having the nucleotide and amino acid sequence selected from the group consisting of the sequences set out in Table 1, and to the nucleotide sequences of the DNA in the deposited strain and amino acid sequences encoded thereby.

Deposited materials

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

A deposit containing a *Streptococcus pneumoniae* bacterial strain has been deposited with the National Collections of Industrial and Marine Bacteria Ltd. (NCIMB), 23 St. Machar Drive, Aberdeen AB2 1RY, Scotland on 11 April 1996 and assigned NCIMB Deposit No. 40794. The *Streptococcus pneumoniae* bacterial strain deposit is referred to herein as "the deposited bacterial strain" or as "the DNA of the deposited bacterial strain."

The deposited material is a bacterial strain that contains the full length FabH DNA, referred to as "NCIMB 40794" upon deposit.

The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

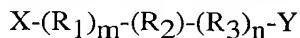
A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The deposited strain contains the full length genes comprising the polynucleotides set forth in Table 1. The sequence of the polynucleotides contained in the deposited strain, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

Polypeptides

The polypeptides of the invention include the polypeptides set forth in Table 1 (in particular the mature polypeptide) as well as polypeptides and fragments, particularly those which have the biological activity of a polypeptide of the invention, and also those which have at least 50%, 60% or 70% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1 or the relevant portion, preferably at least 80% identity to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and more preferably at least 90% similarity (more preferably at least 90% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and still more preferably at least 95% similarity (still more preferably at least 95% identity) to a polypeptide sequence selected from the group consisting of the sequences set out in Table 1, and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

The invention also includes polypeptides of the formula:



wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R₁ and R₃ are any amino acid residue, n is an integer between 1 and 2000, m is an integer between 1 and 2000, and R₂ is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in Table 1. In the formula above R₂ is oriented so that its amino terminal residue is at the left, bound to R₁, and its carboxy terminal residue is at the right, bound to R₃. Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In preferred embodiments n is an integer between 1 and 1000 or 2000.

A fragment is a variant polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned polypeptides. As with polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region, a single larger polypeptide.

Preferred fragments include, for example, truncation polypeptides having a portion of the amino acid sequence of Table 1, or of variants thereof, such as a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus. Degradation forms of the polypeptides of the invention in a host cell, particularly a *Streptococcus pneumoniae*, are also preferred. Further preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil

and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Also preferred are biologically active fragments which are those fragments that mediate activities of polypeptides of the invention, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those fragments that are antigenic or immunogenic in an animal, especially in a human. Particularly preferred are fragments comprising receptors or domains of enzymes that confer a function essential for viability of *Streptococcus pneumoniae* or the ability to initiate, or maintain cause disease in an individual, particularly a human.

Variants that are fragments of the polypeptides of the invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, these variants may be employed as intermediates for producing the full-length polypeptides of the invention.

In addition to the standard single and triple letter representations for amino acids, the term "X" or "Xaa" is also used. "X" and "Xaa" mean that any of the twenty naturally occurring amino acids may appear at such a designated position in the polypeptide sequence.

Polynucleotides

The nucleotide sequences disclosed herein can be obtained by synthetic chemical techniques known in the art or can be obtained from *S. pneumoniae* 0100993 by probing a DNA preparation with probes constructed from the particular sequences disclosed herein. Alternatively, oligonucleotides derived from a disclosed sequence can act as PCR primers in a process of PCR-based cloning of the sequence from a bacterial genomic source. It is recognised that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

To obtain the polynucleotide encoding the protein using the DNA sequence given herein typically a library of clones of chromosomal DNA of *S. pneumoniae* 0100993 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by

Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual, 2nd edition, 1989, Cold Spring Harbor Laboratory (see: Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70).

Moreover, another aspect of the invention relates to isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group consisting of the sequences in Table 1 and polynucleotides closely related thereto and variants thereof.

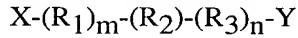
Using the information provided herein, such as the polynucleotide sequences set out in Table 1, a polynucleotide of the invention encoding polypeptide may be obtained using standard cloning and screening methods, such as those for cloning and sequencing chromosomal DNA fragments from bacteria using *Streptococcus pneumoniae* 0100993 cells as starting material, followed by obtaining a full length clone. For example, to obtain a polynucleotide sequence of the invention, such as a sequence set forth in Table 1, typically a library of clones of chromosomal DNA of *Streptococcus pneumoniae* 0100993 in *E.coli* or some other suitable host is probed with a radiolabeled oligonucleotide, preferably a 17-mer or longer, derived from a partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using stringent conditions. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently, such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989). (see in particular Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotides set out in Table 1 were discovered in a DNA library derived from *Streptococcus pneumoniae* 0100993.

The DNA sequences set out in Table 1 each contains at least one open reading frame encoding a protein having at least about the number of amino acid residues set forth in Table 1. The start and stop codons of each open reading frame (herein "ORF") DNA are the first three and the last three nucleotides of each polynucleotide set forth in Table 1.

Certain polynucleotides and polypeptides of the invention are structurally related to known proteins as set forth in Table 1. These proteins exhibit greatest homology to the homologue listed in Table 1 from among the known proteins.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence in Table 1. Also provided by the invention is the coding sequence for the mature polypeptide or a fragment thereof, by itself as well as the coding sequence for the mature polypeptide or a fragment in reading frame with other coding sequence, such as those encoding a leader or secretory sequence, a pre-, or pro- or prepro- protein sequence. The polynucleotide may also contain non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence which encode additional amino acids. For example, a marker sequence that facilitates purification of the fused polypeptide can be encoded. In certain embodiments of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al.*, *Proc. Natl. Acad. Sci., USA* 86: 821-824 (1989), or an HA tag (Wilson *et al.*, *Cell* 37: 767 (1984). Polynucleotides of the invention also include, but are not limited to, polynucleotides comprising a structural gene and its naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:



wherein, at the 5' end of the molecule, X is hydrogen, and at the 3' end of the molecule, Y is hydrogen or a metal, R₁ and R₃ is any nucleic acid residue, n is an integer between 1 and 3000, m is an integer between 1 and 3000, and R₂ is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in Table 1. In the polynucleotide formula above R₂ is oriented so that its 5' end residue is at the left, bound to R₁, and its 3' end residue is at the right, bound to R₃. Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer. In a preferred embodiment n is an integer between 1 and 1000, or 2000 or 3000.

The term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides that include a sequence encoding a polypeptide of the invention, particularly a bacterial polypeptide and more particularly a polypeptide of the *Streptococcus pneumoniae* having an amino acid sequence set out in Table 1. The term also encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or an insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The invention further relates to variants of the polynucleotides described herein that encode for variants of the polypeptide having the deduced amino acid sequence of Table 1. Variants that are fragments of the polynucleotides of the invention may be used to synthesize full-length polynucleotides of the invention.

Further particularly preferred embodiments are polynucleotides encoding polypeptide variants, that have the amino acid sequence of a polypeptide of Table 1 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, that do not alter the properties and activities of such polynucleotide.

Further preferred embodiments of the invention are polynucleotides that are at least 50%, 60% or 70% identical over their entire length to a polynucleotide encoding a polypeptide having the amino acid sequence set out in Table 1, and polynucleotides that are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the deposited strain and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

A preferred embodiment is an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of: a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*.

Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of Table 1.

The invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the invention especially relates to polynucleotides that hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the terms "stringent conditions" and "stringent hybridization conditions" mean hybridization will occur only if there is at least 95% and preferably at least 97% identity between

the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising: 50% formamide, 5x SSC (150mM NaCl, 15mM trisodium citrate), 50 mM sodium phosphate (pH7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/ml denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at about 65°C. Hybridization and wash conditions are well known and exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), particularly Chapter 11 therein.

The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in Table 1 under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said DNA sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers described elsewhere herein.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding a polypeptide and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in Table 1. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of each gene that comprises or is comprised by a polynucleotide set forth in Table 1 may be isolated by screening using a DNA sequence provided in Table 1 to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the invention may be employed, for example, as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

Polynucleotides of the invention that are oligonucleotides derived from the a polynucleotide or polypeptide sequence set forth in Table 1 may be used in the processes herein as described, but preferably for PCR, to determine whether or not the polynucleotides identified herein in whole or in part are transcribed in bacteria in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The invention also provides polynucleotides that may encode a polypeptide that is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in vivo*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In addition to the standard A, G, C, T/U representations for nucleic acid bases, the term "N" is also used. "N" means that any of the four DNA or RNA bases may appear at such a designated position in the DNA or RNA sequence, except it is preferred that N is not a base that when taken in combination with adjacent nucleotide positions, when read in the correct reading frame, would have the effect of generating a premature termination codon in such reading frame.

In sum, a polynucleotide of the invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences that are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Vectors, host cells, expression

The invention also relates to vectors that comprise a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the invention. Introduction of a polynucleotide into the host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*,

(1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), such as, calcium phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host may be used for expression in this regard. The appropriate DNA sequence may be inserted into the expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, (supra).

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

Polypeptides of the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding protein may

be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention is also related to the use of the polynucleotides of the invention for use as diagnostic reagents. Detection of such polynucleotides in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method for diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected with an organism comprising a gene of the invention may be detected at the nucleic acid level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from an infected individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR or other amplification technique prior to analysis. RNA or cDNA may also be used in the same ways. Using amplification, characterization of the species and strain of prokaryote present in an individual, may be made by an analysis of the genotype of the prokaryote gene. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to labeled polynucleotide sequences of the invention. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in the electrophoretic mobility of the DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, e.g., Myers et al., *Science*, 230: 1242 (1985). Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or a chemical cleavage method. See, e.g., Cotton et al., *Proc. Natl. Acad. Sci., USA*, 85: 4397-4401 (1985).

Cells carrying mutations or polymorphisms in the gene of the invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. For example, RT-PCR can be used to detect mutations. It is particularly preferred to used RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to a nucleic acid encoding a polypeptide of the invention can be used to identify and analyze mutations. These primers may be used for, among other things, amplifying a DNA of the invention isolated from a sample derived from an individual. The primers may be used to amplify the gene isolated from an infected individual such that the gene may then be subject to

various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected and used to diagnose infection and to serotype and/or classify the infectious agent.

The invention further provides a process for diagnosing disease, preferably bacterial infections, more preferably infections by *Streptococcus pneumoniae*, and most preferably disease, comprising determining from a sample derived from an individual a increased level of expression of polynucleotide having the sequence of Table 1. Increased or decreased expression of a polynucleotide of the invention can be measured using any one of the methods well known in the art for the quantitation of polynucleotides, such as, for example, amplification, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods.

In addition, a diagnostic assay in accordance with the invention for detecting over-expression of a polypeptide of the invention compared to normal control tissue samples may be used to detect the presence of an infection, for example. Assay techniques that can be used to determine levels of a protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Antibodies

The polypeptides of the invention or variants thereof, or cells expressing them can be used as an immunogen to produce antibodies immunospecific for such polypeptides. "Antibodies" as used herein includes monoclonal and polyclonal antibodies, chimeric, single chain, simianized antibodies and humanized antibodies, as well as Fab fragments, including the products of an Fab immunoglobulin expression library.

Antibodies generated against the polypeptides of the invention can be obtained by administering the polypeptides or epitope-bearing fragments, analogues or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique known in the art that provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies.

Alternatively phage display technology may be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing recognition of a polypeptide of the invention or from naive libraries (McCafferty, J. et al., (1990), *Nature* **348**, 552-554; Marks, J. et al., (1992) *Biotechnology* **10**, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) *Nature* **352**, 624-628).

If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptides to purify the polypeptides by affinity chromatography.

Thus, among others, antibodies against a polypeptide of the invention may be employed to treat disease.

Polypeptide variants include antigenically, epitopically or immunologically equivalent variants that form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably, the antibody or variant thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanized"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for

example as described in Jones, P. et al. (1986), *Nature* 321, 522-525 or Tempest et al.,(1991) *Biotechnology* 9, 266-273.

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., *Hum Mol Genet* 1992, 1:363, Manthorpe et al., *Hum. Gene Ther.* 1963:4, 419), delivery of DNA complexed with specific protein carriers (Wu et al., *J Biol Chem.* 1989: 264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, *PNAS*, 1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda et al., *Science* 1989:243,375), particle bombardment (Tang et al., *Nature* 1992, 356:152, Eisenbraun et al., *DNA Cell Biol* 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al., *PNAS* 1984:81,5849).

Antagonists and agonists - assays and molecules

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics. See, e.g., Coligan et al., *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

The invention also provides a method of screening compounds to identify those which enhance (agonist) or block (antagonist) the action of a polypeptides or polynucleotides of the invention, particularly those compounds that are bacteriostatic and/or bacteriocidal. The method of screening may involve high-throughput techniques. For example, to screen for agonists or antagonists, a synthetic reaction mix, a cellular compartment, such as a membrane, cell envelope or cell wall, or a preparation of any thereof, comprising a polypeptide of the invention and a labeled substrate or ligand of such polypeptide is incubated in the absence or the presence of a candidate molecule that may be an agonist or antagonist of a polypeptide of the invention. The ability of the candidate molecule to agonize or antagonize a polypeptide of the invention is reflected in decreased binding of the labeled ligand or decreased production of product from such substrate. Molecules that bind gratuitously, i.e., without inducing the effects of a polypeptide of the invention are most likely to be good antagonists. Molecules that bind well and increase the rate of product production from substrate are agonists. Detection of the rate or level of production of product from substrate may be enhanced by using a reporter system. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate converted into product, a reporter gene that is responsive to changes in polynucleotide or polypeptide activity, and binding assays known in the art.

Another example of an assay for antagonists of polypeptides of the invention is a competitive assay that combines any such polypeptide and a potential antagonist with a compound which binds such polypeptide, natural substrates or ligands, or substrate or ligand mimetics, under appropriate conditions for a competitive inhibition assay. A polypeptide of the invention can be labeled, such as by radioactivity or a colorimetric compound, such that the number of such polypeptide molecules bound to a binding molecule or converted to product can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polynucleotide or polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing activities induced by a polypeptide of the invention, thereby preventing the action of such polypeptide by excluding it from binding.

Potential antagonists include a small molecule that binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules. Other potential antagonists include antisense molecules (see Okano, *J. Neurochem.* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, FL (1988), for a description of these molecules). Preferred potential antagonists include compounds related to and variants of a polypeptide of the invention.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein, upon expression, can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The invention also provides the use of the polypeptide, polynucleotide or inhibitor of the invention to interfere with the initial physical interaction between a pathogen and mammalian host responsible for sequelae of infection. In particular the molecules of the invention may be used: in the prevention of adhesion of bacteria, in particular gram positive bacteria, to mammalian extracellular matrix proteins on in-dwelling devices or to extracellular matrix proteins in wounds; to block protein-mediated mammalian cell invasion by, for example, initiating phosphorylation of mammalian tyrosine kinases (Rosenshine *et*

al., *Infect. Immun.* 60:2211 (1992); to block bacterial adhesion between mammalian extracellular matrix proteins and bacterial proteins that mediate tissue damage and; to block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

The antagonists and agonists of the invention may be employed, for instance, to inhibit and treat disease.

Helicobacter pylori (herein *H. pylori*) bacteria infect the stomachs of over one-third of the world's population causing stomach cancer, ulcers, and gastritis (International Agency for Research on Cancer (1994) Schistosomes, Liver Flukes and Helicobacter Pylori (International Agency for Research on Cancer, Lyon, France; <http://www.uicc.ch/ecp/ecp2904.htm>). Moreover, the international Agency for Research on Cancer recently recognized a cause-and-effect relationship between *H. pylori* and gastric adenocarcinoma, classifying the bacterium as a Group I (definite) carcinogen. Preferred antimicrobial compounds of the invention found using screens provided by the invention, particularly broad-spectrum antibiotics, should be useful in the treatment of *H. pylori* infection. Such treatment should decrease the advent of *H. pylori*-induced cancers, such as gastrointestinal carcinoma. Such treatment should also cure gastric ulcers and gastritis.

Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with a polypeptide of the invention, or a fragment or variant thereof, adequate to produce antibody and/ or T cell immune response to protect said individual from infection, particularly bacterial infection and most particularly *Streptococcus pneumoniae* infection. Also provided are methods whereby such immunological response slows bacterial replication. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises delivering to such individual a nucleic acid vector to direct expression of a polynucleotide or polypeptide of the invention, or a fragment or a variant thereof, for expressing such polynucleotide or polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response, such as, to produce antibody and/ or T cell immune response, including, for example, cytokine-producing T cells or cytotoxic T cells, to protect said individual from disease, whether that disease is already established within the individual or not. One way of administering the gene is by accelerating it into the desired cells as a coating on particles or otherwise. Such

nucleic acid vector may comprise DNA, RNA, a modified nucleic acid, or a DNA/RNA hybrid.

A further aspect of the invention relates to an immunological composition which, when introduced into an individual capable or having induced within it an immunological response, induces an immunological response in such individual to a polynucleotide of the invention or protein coded therefrom, wherein the composition comprises a recombinant polynucleotide or protein coded therefrom comprising DNA which codes for and expresses an antigen of said polynucleotide or protein coded therefrom. The immunological response may be used therapeutically or prophylactically and may take the form of antibody immunity or cellular immunity such as that arising from CTL or CD4+ T cells.

A polypeptide of the invention or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as lipoprotein D from *Hemophilus influenzae*, Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilize the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

Provided by this invention are compositions, particularly vaccine compositions, and methods comprising the polypeptides or polynucleotides of the invention and immunostimulatory DNA sequences, such as those described in Sato, Y. *et al.* *Science* 273: 352 (1996).

Also, provided by this invention are methods using the described polynucleotide or particular fragments thereof which have been shown to encode non-variable regions of bacterial cell surface proteins in DNA constructs used in such genetic immunization experiments in animal models of infection with *Streptococcus pneumoniae* will be particularly useful for identifying protein epitopes able to provoke a prophylactic or therapeutic immune response. It is believed that this approach will allow for the subsequent preparation of monoclonal antibodies of particular value from the requisite organ of the animal successfully resisting or clearing infection for the development of prophylactic agents or therapeutic treatments of bacterial infection, particularly *Streptococcus pneumoniae* infection, in mammals, particularly humans.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which protect against invasion of bacteria, for example by blocking adherence of bacteria to damaged tissue. Examples of tissue damage include wounds in skin or connective tissue caused, e.g., by mechanical, chemical or thermal damage or by implantation of indwelling devices, or wounds in the mucous membranes, such as the mouth, mammary glands, urethra or vagina.

The invention also includes a vaccine formulation which comprises an immunogenic recombinant protein of the invention together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

While the invention has been described with reference to certain protein, such as, for example, those set forth in Table 1, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions, kits and administration

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or their agonists or antagonists. The polypeptides of the invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration. The

invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device.

Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent bacterial wound infections, especially *Streptococcus pneumoniae* wound infections.

Many orthopedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteremia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1 μ g/ml to 10mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response. A suitable unit dose for vaccination is 0.5-5 microgram/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks. With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

TABLES

Certain pertinent data for preferred polypeptide and polynucleotide embodiments of the invention are summarized in Tables 1 and 2.

Provided in Table 1 are sequence search results providing characterization information regarding certain preferred polynucleotides (denoted as "Assembly") and polypeptides of the invention encoded thereby. For each polynucleotide in Table 1, there is listed the closest homologue of each polypeptide encoded by each ORF in such polynucleotide. This determination of homology is based on a comparison of the sequences

of in Table 1 with sequences available in the public domain (see heading entitled "Description" for the homologue name). Where no significant homologue was detected the term "unknown" appears after the heading "Description". Preferred polypeptides encoded by the ORFs of the invention, particularly full length proteins either obtained using such ORFs or encoded entirely by such ORFs, are ones that have a biological function of the homologue listed, among other functions. The analysis used to determine each homologue listed in Table 1 was either BlastP and/or BlastX and/or MPSearch, each of which is well known. Also provided in Table 1 is the amino acid sequence encoded by each ORF. An "Assembly ID" number provides a convenient way to correlate the polynucleotide sequence with the ORF or ORFs it comprises and the polypeptides encoded by these ORFs, as well as to correlate such sequences with other pertinent information provided in Tables 1 and 2. Following the heading "ORF Predictions" the nucleotides at the beginning and end of the ORF sequence are set forth ("Start" and "End" respectively). The direction of translation on the polynucleotide depicted is denoted by an "F" for forward or an "R" for reverse (reverse being translated on the opposite strand from the one depicted). The length of each amino acid sequence is also indicated in a column entitled "Length." Below these data is shown the amino acid sequence encoded by the ORF. If a given polynucleotide comprises one ORF, then in the column entitled "ORF #" there is the numeral one. If it encodes two, there are the numerals one and two in the column, and so on.

TABLE 1

Assembly ID: 3047950

Assembly Length: 587bp

[SEQ ID NO:] 3047950 Strep Assembly -- Assembly
id#3047950
CTCAGTTCTGCCATCCTCTCCTCGCTTTTGATGAAACTGCCCTTCATATCTACAC
GCTTGTCCAGATAGCGATAAACCGCGCTGATATCCATCTCCATGAAATAGGTTGGGGCAA
ACAGTTGATTTAAAATGTCCCTTTCATCCAGGAATTCTGGGGCAACAAGTCGCTCAA
GAATCTTGGCAAAGATGTGGCAAATACCGTCTCCTCAACAATCCTATCTACCCGACAAT
CTAAAACAAGTGGACAGGGCGTCTAAAATAGAAATCTGAGTTGTTCAAGAAATTTCATAAT
GCACTCCAAACGTTCCAATTCTCCTGATGACTGATAAAACCAGCCTGCTCCATCGCAA
GCATAGAAGTTCATCAGAAATATTCACAGTAAATTGGATACTGTTGATCTGCTCTG
CGGCATTCTCTCGCAACGACTCCAATCACAACCCAATCTCCTAGACTATAAGAAGAAC
TACAGGTCGTGATGTTAGCCAAAATTCTAATCTGATATCCTAAAATAAAACAGGAA
AACCATATAAGTTACTTGTGTTAAAGATTGCTTCATAACAACC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	2	451	R	150 aa

[SEQ ID NO: 3047950-6] 3047950-6 ORF translation from 2-451, direction R

VIGVVARENAAEQIKQYQKFTVNISDETSMLAMEQAGFISHQEKLRLGVHYEISERTQI
SILDACPLVLDCRVDRIEEDGICHIFAKILERLVAPEFLDEKGHFKNQLFAPTYFMGDG
YQRVYRYLDKRVDMKGSIKKARKKDGGN*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3049152

Assembly Length: 468bp

[SEQ ID NO: 3049152] 3049152 Strep Assembly -- Assembly id#3049152

CTTCCTAGTTGCTTTGATTTCATTGACTATAATGGTTTAATTCTTTTTCAAATCTGGCACTACTTCTGCCTCAAACCAAGGATTTGGCCATCCAGATTGATTTCGTGGT
GATGGGTGAACTAGCGAAAATAGGCTGGCAGATAGTCTTATAGTGTTCACCCCTCTCC
GTTACCTTCCCACTGATTTCTCCTGTAAATAGTAGGCTTGGCATATTGCCAATCAAG
AGGGTTAACTGAATATCAGGCAATTCTGTAAAGAGCTGCGGATGCCATTCTGCAAAA
CCTGTACGAGGCGGAAGATCACCCGACTGCCATGTCCTGGAAAGTTAGAAATCCATAGG
CAAAACAGCAAAATAACCTGAATTGTAAAAGGTATCTCATCCACACCTAGCCAGTCCCC
GCAAGCGGTACCACCTTATCTTCCAGTAAGCCTGCTTCCTTGATT

ORF Predictions:

ORF #	Start	End	Direction	Length
-------	-------	-----	-----------	--------

6 24 407 R 128 aa

[SEQ ID NO: 3049152-6] 3049152-6 ORF translation from 24-407,
direction R
VWMKIPFTIQVILLFCLWISNFPGHGKSGDLPPRTGFAEKWHPQLLQELPDIQLTLLIGQ
YAQAYYLQEKGKVTERVKHYKDYLPAYFPLVHPSPRNQIWMAKNPWFEAEVVPDLKRR
IKTIYSQ*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3174820
Assembly Length: 1086bp

[SEQ ID NO: 3174820] 3174820 Strep Assembly -- Assembly
id#3174820
CTACCTTGCTAGATGTGATAGACCGTGGGAATGTCTCTATCATTTCAGAAGGGAGATGCAG
TTGGTTTGAGGCTAGTAAAAGAAGATGGTTGTCAAGCTTGAGAAAGACTGCCTAAATC
TAGCTTTTCAGGTAAGAAGAAACTCTTCCAATTGTTGCGGATTACAAGGTAT
CTGATAGTCTTATCGTAGAGCCAAAGTTCTGATGAAAACGGATTCAAGCAAGAGGGC
TTCAACTCAAATCTTCTTGAAGAGGTATTGAACCAGATGCAAGAAGGAGTGAGAAAAC
GAGTTCCCTCTGGGGCTCCAGATTACTATCGTCCTTAACGGTTGGAAAAGGCTT
TGCAAGTGGGTATGGGTCTTGACTATCTGCCCTATTATCGGATTGGTTGTTCT
TGTACAGTTAGACGTTCATGGCTATCTTACCTCCCTTGCCAATACTGGTTTCTAG
GGTAGTTTGTCTGTTCTATTATTGGAAGCTCGACTAGATAATCGTATGGTGTTC
TAAATGAAGCGGGAGCTGAGGTCTACTATCTGGACCAGTTGAAAATATGTTACGTG
AGATTGCACGACTGGATAAGGCTGAATTGCGAAAGTATTGTTGGAATCGTCTCTTG
GTCTATGCAACCTTATTGGCTATGCGGACAAGGTTAGTCATTGATGAAGGTTCATCAG
ATTCAAGTTGAAAATCCAGATATCAATCTCTATGTAGCTTATGGCTGGCACAGTATGTT
TATCATTCAAGCGCGAAATGAGCCATTATGCTAGTGTGCAAATACAGCAAGTACCTAC
TCCGTATCTTGGAAAGTGGAAAGTCTGGTGGCTTCTGGAGGCGGAGGTGGCGGCA
GTATCGGTGCCTTTAAAGAGAGCTACCACACTGAAAAGTATGATATATGGAAGATA
GAAAAAGACACCTATANGAAAATCATAGTTTATCTAAACTATTCTTATTCCATTGAT
GATTTGGCGAAGAATTAGAACCCGGAAAAAGCCCTTGAAAAATTCCATTTCCTAA

AGGTAA

ORF Predictions:

ORF #	Start	End	Direction	Length
7	598	1041	F	148 aa

[SEQ ID NO:] 3174820-7 ORF translation from 598-1041, direction F

VRLHDWIRLNCESVVWNRLLVYATLFGYADKVSHLMKVHQIQVENPDINLYVAYGWHSM
 FYHSSAQMSHYASVANTASTYSVSSGSGSLVVASLEAEVAAVSVPFKESYHTLKKYDIWK
 IEKDTYXKIIIVLSKFLISIDDFGEEF*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3175500

Assembly Length: 1284bp

[SEQ ID NO:] 3175500 Strep Assembly -- Assembly
 id#3175500

CTCATTGCAAAATCAGGAAAAACGGATGGTAACGGCAGTCGAAATGTTCTATCTAAGA
 AACAAAGAGGCTTGAAAAAAATGCAGTCAGTCTGTTATCTTAGACAACCTGAGCGCTTG
 ATGACGGTTATTTGCAACGCTTGGACCAACTGCAACTGCGTTGAAACAAAGTTGC
 GAACTCGGATTCTGATAACAAACAATTAGTTCAAGCAAGAACTCATCAATTAGTACA
 ATTACACCTGTTACCAAAATCCAACGCTATCAAGACCGTTAGGACAGTTGGACAAG
 CGCTTCTTA GGTAGCCAATGGCGTTAGTTATGACGCCAAGGTTGCTGAGGCCAAG
 CGACTTTGGAACTGGCTCATGTTGGATACTAGGCCAATCGTGGCGTGGTTATGCT
 ATTGTCAAAAAA GAAGAATCCGTTGAGATTGGTTGAGAGTTGAAGAAAAAGACCAAG
 TAACGCTTTG ATGCGAGATGGTCAAGTAGAATTAGAGGTAAAGATGTAAA
 ACAAAAGAAATTGAGGA AAATCTAGCAGAACTGGAAACCATTGT
 CCAAAGTTGGAAATGGTGAATTGCTCTGGAAAGATGCGATTACTGC
 CTTCAAAAGGGCATGGTCTGTCAAAAGAGCTCCAAGCTACGCT
 GGACAAGGCTGAAAAGACCTTGGTCAAGGTCAAGGTCAAGAAGAC
 GGAACAGAAAGTGATTT

TGAATGAAAAAGCAAGAAAAATTAGCTCTGTCGAGTCGGCTTGGAAAGATTTATGGAG
 ACCAGCAGTTGCCTCTAGTTACGGGAGTCTGTTCTATTCTATTGCTGGTGGCA
 AGCGTATTGCGCTTTCTCTGTTAGAAGTTCTGGAAGCCTGAGGTTACCATCAAAC
 CTGCTCNCGCGCAGGTAGCTACTGCCTGGAGATGATTACAGGGAGCTGATTACCG
 ATGACCTTCCTGCTATGGATGATGACGAGGATCGAGAGAGGGCGGAAAAACCAATCACAA
 GAAATCCGGTGAAGCTATGCCATCCTAGCTGGAGATGCCTCATGCTTAGACCCATATGC
 CTTGATTGCGCAGGCAGATCCGCCAAGTCAGATCAAGGTGGGCTCGATTGCCAACTCATC
 CCTTGCTTCAGGTAGCCTGGGTATGGTGGCAGGGCAAGTCTGGATATGGAGGGCGAACAA
 CCAGCACTGGTCTCTGGAAGAACCTCAGACTATGCATGCCAACAAAGACTGGGAAGTTACT
 AGCCTATCCCTCCAACGCGGCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
8	714	1049	F	112 aa

[SEQ ID NO:] 3175500-8 ORF translation from 714-1049,
 direction F
 VILNEKARKISSCRVGFGRFYGDQQFASSLRESVLYSIHAGGKRIRPFLLLEVLEALQVT
 IKPAXAQVATALEMIHTGSLIHDDLPAMDDDEDRERAEKPITRNPVKLWPS*

Blastp and/or MPSearch Result:

Description:

GERANYLTRANSTRAINSFERASE (EC 2.5.1.10) (FARNESYL-DIPHOSPHATE
 SYNTHASE) (FPP SYNT HASE). - BACILLUS STEAROTHERMOPHILUS.

Assembly ID: 3175674
 Assembly Length: 816bp

[SEQ ID NO:] 3175674 Strep Assembly -- Assembly
 id#3175674
 CTGTTGGAAAATAGGTGCTTTAAATTGCCAGTAGAAGTGGTTCAGTATGGTGCAGAGC
 AGTCTTTCGTCAATTGAAACGAGCTGGTACCAAACAAGTTCCGTGAAAAGACGCCAAC
 GTTTTGTGACGGATATGCAGAATTTATCATTGACCTCGCCTGGATGTCATTGAAAATC

CAATTGCTTTGGACAAGAATTGGACCATGTCGTTGGTGTGGAGCATGGTTATTCA
 ACCAAATGGTGGATAAGGTAATCGTTGCTGGACGAGATGGAGTCAGATTCAACTCAA
 AAAAAGGAAAATAGAAGGGGGCATAAGATGTCTAAATTAAATCGTATTCAATTGGTGGTA
 CTGGATTCTGTAGGAATCGGTGCAGCACCAGATGCTAATAACTTGTCAATGCAGGGTT
 CCAGATGGAGCTCTGACACACTGGGACACATTCAAAAACAGTTGGTTGAATGTCCCA
 AACATGGCTAAAATAGGTCTTGAAATATTCCCTCGAAACTCCTCTTAAGACTGTAGCA
 GCTGAAAGCAATCCAACCTGGATATGCAACAAAATTAGAGGAAGTATCTCTGGTAAGGAT
 ACTATGACTGGACACTGGGAAATCATGGACTCAACATTACTGAGCCTTCGATACTTTC
 TGGAACGGATTCCCAGAAGAAATCCTGACAAAATCGAAGAATTCTCAGGACGCAAGGTT
 ATTCGTGAAGCCAACAAACCTATTAGGAACGGCTGTTATCGATGATTGGACCACGT
 CAGATGGAAACTGGAGAGTTGATATCTATACTTCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	126	314	F	63 aa

[SEQ ID NO:] 3175674-6 ORF translation from 126-314,
 direction F
 VTDMQNFIIDLALDVIENPIAFQELDHVVGVVEHGLFNQMVDKVIVAGRDPVQISTSKK
 GK*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3176442
 Assembly Length: 617bp

[SEQ ID NO:] 3176442 Strep Assembly -- Assembly
 id#3176442
 CTAGTACAGCTTATGCGGCCGTTTATTCCGAACATCCAGATCAGCCCTTGCAGCAA
 TTGCACCCAGAATTCTGCTGAAGAATATGGATTGGAACCTGATTGCCGAGGATATTCAAGG
 AAATGGAAGCCAATTACACAGTTCTGGCTCTAGGAGCTGAAAAGCCTAGTATTCCCT

TGCAAGCACAAACTGAAAAGATGAGTTGGCCTTGACATTACCTGACAACCTTCCAGGTG
 CACTTATAAGGCCCTGTCGACCTTGCTTGGCGAAGGGAATTGACTGACAAAAATTGA
 AAGTCGTCCACTCAAGACAGCACTGGGTGAATACTTTTCATTATCGATGTGGATTATAC
 CGATAAGGACTTGGTCCACTTGCCCCAAAAGAATTAGAAGCGATTGGAATCCAGTATAA
 AATTCTGGGTGCCTATCCTATTATCCAATATCAGACCATGGAAAGGAGAGAAGATGAGT
 AAAGAAAATCCCTTAAGTCATCATGAGCAGTTGCCTATGATTATTGCTAAAAAATT
 CACTATCTCAATGAGAGAGAAAAAATGAGTTGTCTATTGCAAGAAAAGCTAACTCTT
 GCTAGGGGAAATAGTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	350	478	F	43 aa

[SEQ ID NO:] 3176442-6 ORF translation from 350-478,
 direction F
 VDYTDKDLVHFAQKELEAIGIQYKILGAYPIYPISDHGKERR*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3176630

Assembly Length: 457bp

[SEQ ID NO:] 3176630 Strep Assembly -- Assembly
 id#3176630
 CCAGTCATCAAATTGACCAAATTGAGAGTCAAATTACTTGATTGAAAAAAATTGCGG
 CAATTCGCAATGCTTGGCAGACTTAGAGAAGCAAGAATCTAAAATAGTGGTCGTGTT
 TTCATGCTTGGATTATTGAGGAACCTCAGCATAAGTTGCTGAAAATTCAAGAACAGT
 ATGGTCAAGCCTTGGATGAAATTGAAAAACAATGAGAAAATATCCAATCTGAATTTCAC
 AATTGTAAACCTTGAATTTCATGGGTGACCCCTGTGGAAGGCCGAGTGATTTGGATAATA
 CAGAAAATCACATTGGCCTTAAGTCATATTGTGGATCGTGTCCAGCCTGGTTACGA
 CCTTCTACAGAATTGCCAGATCAATTACAGGGATTGGAACCGGTTATCGTAAACTAAT

TGATGCTAATTATCATTGTTGAAACGGATATGGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	273	419	F	49 aa

[SEQ ID NO:] 3176630-6 ORF translation from 273-419,
direction F
VEAAVILDNTENHILALSHIVDRVPALVTTFLQNCQINYRDLEPVIVN*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3176662
Assembly Length: 381bp

[SEQ ID NO:] 3176662 Strep Assembly -- Assembly
id#3176662
CTTATTAGTACGCATTCCCCTGTGGGAAGTAAGTTCTCTGGCATGTCGTTGATGA
TGACATGGACAGCAGATTGAGGGGCTCCAGTGGTGCACACTGCTCCGTTACTTCCT
TAGCAAGAGCTTCTTGCTCGAGCGTGCCTCAAATAATCGATGCGTACAAATG
GCATAATAGCTCCTCCACTAGTTGATTCTCCATTACACATTGCGTTAA
AGCTTAAGAAAATTATGATATACTAGAAATGTAGCAAAATTAGAAATGGACGTGAAGCA
AGAAACATGGCACAGTTGACTATCGTTATGGGACCATGAACCTCTGGTAAAACGATTGAG
ATTCTCAAAGTGGCCTATAAC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	2	226	R	75 aa

[SEQ ID NO:] 3176662-6 ORF translation from 2-226,
direction R
VVWKWKS KLVEE AIMPFV RIDLF EGRTL E QKK ALAKE VTEA VVRNTGAPQSAVHVIINDM
PEGTYFPQGEMRTK*

Blastp and/or MPSearch Result:

Description:

4-OXALOCROTONATE TAUTOMERASE (EC 5.3.2.-). - PSEUDOMONAS
PUTIDA.

Assembly ID: 3857692
Assembly Length: 743bp

[SEQ ID NO:] 3857692 Strep Assembly -- Assembly
id#3857692
CTGGCAAATA CAAAGGTGACGATCATTGGTAAATCAGCCCACGGTGCTATGCCTGCTTCAG
GTGTCAATGGTGC GACTTACCTAGCCCTT CTTAGCCAGTTGACTTGCTGGTCCAG
CCAAGAAATA CTTGACATCACTGGTAAAATTCTCTTGAAACGACC ATGAGGGTGAAAGTC
TCAAGATTGCTCATGTGGATGAAAAGATGGGTGCCCTTCTATGAATGCAGGCGTCTTCC
GCTTCGATGAAAACAAGTGCTGATAATACCATTGCCCTCAACATCCGCTATCCAAAAGGAA
CAAGTCCAGAACAAATCAGTCAATCCTGAAAAC TTGCCAGTTGTTCTGTTAGCCTGTC
TGAACACGGTCACACGCC TCACTATGTGCCAATGGAAGATCCACTTGTGCAAACCTTGT
GAATGTCTATGAAAACAAACAGGCCTTAAAGGTCA TGAACAAGTCATCGGTGGTGGAAAC
CTTTGGTCGCTTGT TAGAGCGGGAGTTGCCTATGGTGCTATGTTCCCAGACTCAATTGA
TACCATGCACCAAGCCAATGAATT TATTGCCCTGGATGATCTCTCCGAGCAGCAGCAAT
TTATGCCGAAGCTATTACGAATTGATCAAATAAAAGATAGAAGTCTGAGATCTTATGC
TTGGACTTCTTTGGAGGGAAAGTAGATGTCTCAAATCGAAAGAATCAAACAGGCTATC
ATGGCGGATTACAGAATGCCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	386	634	F	83 aa

[SEQ ID NO:] 3857692-6 ORF translation from 386-634,
direction F
VPMEDPLVQTLLNVYEKQTGLKGHEQVIGGGTFGRLLERGVAYGAMFPDSIDTMHQANE
IALDDLFRAAAIIYAEAIYELIK*

Blastp and/or MPSearch Result:

Description:

XAA-HIS DIPEPTIDASE (EC 3.4.13.3) (X-HIS DIPEPTIDASE)
(AMINOACYL- HISTIDINE DIPEPTIDASE) (CARNOSINASE). -
LACTOBACILLUS DELBRUECKII (SUBSP. LACTIS). (BLAST)

Assembly ID: 3857944

Assembly Length: 1783bp

[SEQ ID NO:] 3857944 Strep Assembly -- Assembly
id#3857944
CCACGGTGGAGGGTTGCAAAGTAAGCGACGAATTGCGTTGGTACGACCATTGAAATTGGT
GAGAGGTATGGATGTACGGTCGTAAGGACGATATCGTCGGTATCTTGGCTACATTCTCT
TCTACGATAGTGAGGACTTGGCACCAACGGGCTGCGACCTCTGGATATTCCACGAGTA
TGGTGGCAAGAACTGGATCTGACAAGAGAGCCAAACAGGCCTTCCTCAATCAAG
GCAATGGTCCGTGCTTGAGTTCTCCTGCTGCAAAACCTCACACTGGATATAAGAAATC
TCTTGAGTTGAGACTTGCTTCCATGGCTACGTAGTAATCTGACCACGTCCGATGTA
AAGGCCTACGAGTTCAAGAAAGTCCACGAACCTTGACTTCAATGGTTCTTCTCT
GAAAGAGTTGATTCCAATAGACTGAGCTACGATTGACAATTGACCAACCAGGTCAAAGGC
TTGCGCTTACGATTACCATTTGCTTCTCCACTGCTTTGCAAGGAAGGCAAGGGCTGC
GATTTGCGCTGTATAGGCTTCTGATGCCACGGAATTCAAGGACCTGCGTGAAGGAG
CATGGTATAGTTGGCTTCACGTGAGAGGGTTGAACCTGGAACATTGTCAGTGTAAAGCT
TGGAATTCCCATTCATTGCTTGACCAAAACTTGACGACTATCCGCTGTTACCAGA
TTGGCTGATAAAGATGAAGAGTGTTCTGCTGAGAAGTGGCATAACCGTAGCCCCACTC
AGATGAAATTCCAAGTTCAACTGGTGTATCTGTCAATTCTCAACATTGTCAGTAAAGC
AAATCCTGCATGGTAAGATGTTCCAGCTGCAAGGATGTAGATGCGGTCTGCGTCTGAAC
AGCCTTAATGATAGCAGGATCAACCACTACTTGACCAAGCAGCATCCGTGTTAGGCTTGAAT
GAGTTACGCATAACAGTTGGTTGCTCATCAATTCCCTTAAGCATGTAGTAAGGATAAGT
TCCCTTACCGATATCTGACAAGTCAAGTTCCGCACTAGCTAGCACGTTCACGACTGTT
ACCATCATAGTCTGGAACTTCCACGCTATCAGCCTTGACGATTACCAACTCTGGTCAT

GGATTTCCATGTATTGGTTAGTTCACGAATCATGCCATGGCGTCTGAGCAGACCATGT
 TATAGCCTCTCCAAGACCAATCAAAGTGGTGATTATTAGCTACGTAGATGACTT
 CAGGATCTTGTGAGTCAACCAAGGCAAAGGCATAAGAACACGGATGATGTGAAGGGCTT
 TTTTGAAGGCTTCAAGAACTGAGAGCCCTTCTTCCGGCAAATTCCAATCAAATGA
 ACGGCATTTCAAGTATCTGTCTGCCCTTGAAGTGGTGACCTGCAAGGTATTCTCCTT
 ATTTCAAGATAGTTCTCAATCACCCATTATGCACCAAGACAAAACGTTCTGTCTCAGAG
 CGGTGTGGGTGAGCATTGTCTCAGTTGGTTCCGTGAGTAGCCAACGAGTATGTCCG
 ATACCAAGTTGTTCCCTCAACACCCGGCTGTCTGGCAGACAAATCGATGCAATACGACCAA
 CCGCCTTCACCAAATGGTTATCAGCACCATTAGGACAAAATTCCCGAGAATCATAGC
 CACGGTATTCAAGCTTTCAAGCCCTGAATCAAATATCAGTTGCATTGTGTTCCAA
 CAACACCAACAATTCCACACATAGTATACGACACAGGCAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	1332	1475	R	48 aa

[SEQ ID NO: 3857944-7] 3857944-7 ORF translation from 1332-1475, direction R
 VHNGVIENYLEIKEEYLAGHHFKGQTDTEIAVHLIGKFAGRRRALSS*

Blastp and/or MPSearch Result:

Description:

PROBABLE GLUCOSAMINE--FRUCTOSE-6-PHOSPHATE AMINOTRANSFERASE (ISOMERIZING) (EC 2.6.1.16) BSU21932 NCBI gi: 726479 - Bacillus subtilis.

Assembly ID: 3858118

Assembly Length: 1729bp

[SEQ ID NO: 3858118] 3858118 Strep Assembly -- Assembly
 id#3858118
 CTCAGCTACTTCGCCTTCTTTTATTCTACTGGTTTTCTGATTCCAGTAGTTGAG
 AAGATTCTGTTTTATTCTGAAGTTGATTCAAGCTTCTGATTGCTTGTATTGC

TTGGTTGTTCGTCGCTAGCAGTTCAATGTTAGATTCTGCAGTTGCAGTTGGTTGGT
 TCTCAGCACTGGTGTATCACCATTTGCTCAGCATTGCTGGACTGTTCTTCAC
 TTGCGCTAGCTTGACTGGATTGATGATTCAAAACTAGAATAGCTTTGTCGATTCAA
 GTAAAGCTGTTGTCTTACTATTAGCAGAAAGTTGATCTAATAATGCATCCACCTAT
 CAAAAGTCCGCATCAGATCCATTATTACTTCTAAATAAAAGTGAAGCGACATGAGAATA
 TCGTAGAGTTTGATAGAGTACAAGTGTCTGAGGATCTGCTCAGCATTTCCTTCT
 TGTTGAAGGGCGCTAGCGATACGAGTCAAGACATCTTACCTGACTGTTACTTCATCC
 AAGTCTGCATCAGCCTGTTGTCAGCTTTAGATTTCTACTTCTTGCCAAAGAT
 TGTCTGATTCCCTCTCATGGATCGTCCAAGAGTTGATTGCCTGCTCAAAAGACTT
 TCTACTTCTCCTGCTATCTGTCGAGATTATTGGTTGCTATCTACCATGTAACCTCAA
 AACAGGAGAGTTATAATCCAAGATTACAAGGCCTTACAGAAATAAGAAATCCAGATAAGA
 CAATGTTCGTCCAAGACGCTATTGCTCGCACAGCAGCACGGATTCAATATGCTTTAAT
 TTTAAAGTTAGGTGTCAAGACCTCTTTAGTGTGCCAAAATTAGAGAAGTAATCAA
 TCAACTAACTTTATTTTCAAACTTCAAGTAACTGACCTAAAGCTAACTCAATCTG
 TCTTGTGATAGGCTGTCTTGATAGTGTCTGCTATCAGATCTAGAAGTTGATCT
 ACTTTGCCAAGACTGCCTCTCATCAAAGTCCAGGTTGATAGTTGGATTGCAGGGAT
 GGAATCTTGTCTTCAAAGCCGCTTCATATCCCTAGTTGAACCTTGATGTAGTGTGATTG
 TGGTCGCCACGAGGAATCACAAACCTCTGAATCTCACTTATAATTGATTGGCATCA
 AAACCATGACCCTCTTCTCATGGTGACATGTAGTGACGGATTACTTAATACAGAA
 CTAGAAGAACTTCCTACCTTCCGTAGAGTGTGATGGGGATTGTTAAGAGATGAC
 TTAGGAATATAGTGTAGTGACCCCCATGTCTTACTATATAAGCATCACCTGTATCTGA
 CAATATCATTAGGGTAAAGACATAACCACATCTGCTGCAGAAACACCATTATCGGTG
 TCACCGACAAAGATTGACTGAGAGCTGTAGTATTCTCTGATAATTATACTTTGCAGCTG
 CTAATTCACCTGCCACAAGTCACTCTCAGGAATGAAATGATAGTGACCAACATGTGGTA
 CTATAGTAGATTGAAATAGAATATGAGCAAATTGATAAGGGATTAAAGTAATTCTA
 ACAATGATTAGAAACTATGATGTGCTATTCTAAATTCAACTCACTATATAACCACATCA
 TCGGTAGTATAACGTCCCTGTAATTGCTACAGATACTTCTGCACTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	948	1160	R	71 aa

[SEQ ID NO: 3858118-7] ORF translation from 948-1160,
 direction R
 VIPRGDHNYIKVQTKGYEAALKNKIPSLQSNYQPGTFDEKAVLAKVDQLLDLIAESTK
 TSLSNKDRLS*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3858152
Assembly Length: 1047bp

[SEQ ID NO:] 3858152 Strep Assembly -- Assembly
id#3858152
ATATTCTCAACC ACTGGAGATGGCGCTCGATATCCATGATTAGATTGCGAACGAAAAGAC
GGGTCAGCTCCAGCTGGCTTCACCAGGACCACGGGAACCAATTCCCCCTGCCTGACGGC
TGAGCATAATCCCCTGACCAACCAAGCGAGGCAGAGGTATTTGAGTGGGCTAGGTGGA
CTTGGAGCTCCCTTCATGGCTTCAGCGAGCCGATGGCAAAGATATCCAAAATCAACTGCA
TACGGTCAATGACCTTAACACCGAGAACTCCTCTAGATTGACATTCTGCCTGGGTCA
GACGGTTGTTGACGATGACAGTAGTGATTCTTCTGCATCCACCAAGCGCAATCTCTT
CCAACCTTACCAAGAGCCGACGAAGGTCTTCCAATCATATTTTACGTTTGTCTGTAGC
TATCTACAACCGACTGCCCTGCCGTTTCGCTAAACTAGCCAATTCTCCATGGAGAGGT
CAAACACTGTCCATACCCCTGCAATTCCACACCAATCAGCAGGACTCGCTCCTCTTTCT
CCGTTCAATCATCTAAAAACTCCTCTATCTGGCTTAAATGCGGTCTGTACACCAGAT
TCTCCAATCTGATAAAAGGTGACCTGCATGCGATTACGGAACCAGGTCACTGACGCTTG
GCAAAACGACGGGTCGCCTGTTAAGACTCTCACGAGCTCCTCAAAGGTCTGCTCTCCA
CGGAAATAAGGAAAGAGTTCCATTAGCCAATTCCCTTAGCAGCCTGTACATTAGGGAA
TGGTCAAACAGCCACTTGGCCTCATCCAAAAGCCCAGCCTCAAACATCAAATCCACTCGG
TGGTTGATACGCTCATAAAGTTGACTACGTTCATCCAAAGCAGATAATCAGCGGTCA
TACAAGATCTCTGATTTCAAATCCTGACCAAAATGGCAATTGATGGCACCGCATAG
CACGACGACGATTAAACTGGGAATCTCAAGGCCTGCTTGCTCCACCAAATGGCTAATT
CCTCATCTGAATATGGCTCCAAATTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	546	836	R	97 aa

[SEQ ID NO:] 3858152-6 ORF translation from 546-836,
direction R
VDLMFEAGLLDEAKWLFHDSPNVQAAKGIGYKELFPYFRGEQTFFEEARESLKQATRRFAK

RQLTWFRNRMQVTFYQIGESGVQDRILSQIEFLDD*

Blastp and/or MPSearch Result:

Description:

TRNA DELTA(2)-ISOPENTENYL PYROPHOSPHATE TRANSFERASE (EC 2.5.1.8) (IPP TRANSFERAS E). - AGROBACTERIUM TUMEFACIENS.

Assembly ID: 3858258

Assembly Length: 1565bp

[SEQ ID NO:] 3858258 Strep Assembly -- Assembly
id#3858258
TCGAATCTGGATATGGAGATTGCCAACCATGTCGTGGCTTGGGGCAAGGAAATCGAT
GTTCTGGAAAATCTGACAGTCGCTGGAAATTAAAGCAAAGAGCTGCCAGTCGGAAGT
TTCTATTGTCAACCAAGAACGAGAACAGGAAATCAAGGACTATATTGACCAAATCAAAC
GTGATGGTGATACCACCGTGGGTTGGAGACAGTCGTCGGAGGCGTCCAGTTGGTC
TTGGTTCCTATGTCCAATGGGATAGAAAATTGGATGCAAGATTGGCTCAAGCTGTTGTCT
CTATCAATGCCCTTAAAGGGTGGATTGGCTTGGCTTGAGGCTGGTATCGTAAAG
GCAGCCAAGTATGGATGAAATTCTCTGGCTAAAGAACGCGTTACTCGCCGTACCA
ATAATCTAGGTGGTTTGAAGGTGGTATGACTAATGGCAACCCATCGTGTGGCTGGG
TCATGAAACCCATTCTACTCTTATAAACCTCTTATGAGTGTGGATATCGAAACCCACG
AACCTTACAAGGCAACCGTGGAGAGAAGTGATCCGACTGCTCTCCAGCTGCAGGAATGG
TCATGGAAGCAGTTGTAGCAACGGTCTGGCGCAAGAAATCCTCGAAAATTCTCATCAG
ATAATCTTGAGGAACTAAAAGAACGCGGTAGCCAAACACCAGAGACTATACAAAGAACTATT
AAGGAGTTCCATGGCAAAACAATCTATATCGCAGGTCTGGGTTGATTGGAGCCTCTA
TGGCACTTGGTATCAAACCGCATCCAGATTATGAAATTAGGTTATAATCGTAGTC
AAGCTTCGAGAGATATCGCCTGAAAGAACGGCATGATTGACCGTGCAACGGATGATTG
CTAGTTTGCTCCTTGGCAGATGTCATTATCCTCAGCTTGGCAATCAAACAAACTATTG
CTTTCATTAAGGAGTTGGCCAATTGGATTGGCAGAACGGCTTATTATTCAGATGCTG
GTTCGACCAAGTCAACCATGGATGCGCGGAGCAGTATTGGCTGGCAAGTCTGTT
GCTTGTGGGCCATCCATGGCTGGTAGTCACAAGAACAGGGCTGCTCGCAGATG
TCAATCTTTGAAAATGCCTATTATATCTTACACCTCAAGCCTGACAAGTCAGGACA
CGCTTAAGGAAATGAAGGATCTGCTTCAGGTCTTCAGTCTCGTTATCGAGATTGATG
CCAAGGAGCATGATCGTCACCTCTCAGATTAGCCATTTCTCATATTTGGCTTCTA
GTCTCATGGAGCAGACTGCGGTCTATGCTCAAGAGCATGAGATGGCAAGGCGCTTGC
CAGGTGGTTTCGAGATATGACCCGAATTGGCAAAGCGAGGCCAGGAATGTGGACCTCCA
TTCTCTGTCCAATAGCGAGACCATTCTGGATAGAATTCAAGGATTCAAGGAACGTTGG

AAGCGATTGGTCAGGCCATTAGTAAGGGAGATGAAGAGCAAATTGGAACCTTTTAACC
AAGCG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	207	722	F	172 aa

[SEQ ID NO:] 3858258-6 ORF translation from 207-722,
direction F
VETVVGGVPVGLGSYVQWDRKLDARLAQAVVSINAFKGVEFLGFEAGYRKGSQVMDEIL
WSKEDGYTRRTNNLGGFEGGMTNGQPIVVRGVMKPIPTLYKPLMSVDIETHEPYKATVER
SDPTALPAAGMVMEAVVATVLAQEILEKFSSDNLEELKEAVAKHRDYTKNY*

Blastp and/or MPSearch Result:

Description:

PHOSPHO-2-DEHYDRO-3-DEOXYHEPTONATE ALDOLASE, TYR-SENSITIVE
(EC 4.1.2.15) (PHOSP HO-2-KETO-3-DEOXYHEPTONATE ALDOLASE)
(DAHP SYNTHETASE) (3-DEOXY-D-ARABINO-HEP TULOSONATE 7-
PHOSPHATE SYNTHASE). - BACILLUS SUBTILIS.

Assembly ID: 3858314

Assembly Length: 983bp

[SEQ ID NO:] 3858314 Strep Assembly -- Assembly
id#3858314
CTGATTAGTTTCCTCTTTTGTGTTTCAAACCTAGACCACCGAGTAAACCTGCAAGCG
CAAGCCCAAGGAAACCAATACTGCCATTGATGTTGAGTCTCACCAAGTATTGGTAGCA
TAGCTTATCCTCTGACATCATCGTATCAGACATCTGTTAGCAGAACGCCATGTTT
CACCTGCCATCGTGTGTTAGAACCTGTCATGGTGTCAAGCAGGCTATCTGTAATAC
CTGTAGCATGATTGTGATTCATCGGAGTCACGCCAGAACCCAGAGTTAGAAGGTGATAATG
AACCATTTGCTGTCTGAAGTTCTTAACATTATCTTAATAGTGACTTTTTAGTTG
CTACGATGTTGTCCAAGTCTGGTTACCGTCTTGTACCATAGACATTGACTGTAGCGC
TGTAAGTTGAGTACCAATTGCTCGGAACCTGGTCAATGAGCGCTTGTGTTTCTTGCCAG

CTACATTCCGTCCAAGGCTACTTGATAGAAGTATTGACCTTGGTCTCACGTTTCAC
 CTAGTGGAGATAGGGCTGGGTTTTAGCGTCGCCGTTATCTGACCATGGTGCCTTGTCA
 ATGCCTTGAGCAAGAGACGAGTCAACATACCATCACCTGCGAAGAGAGTTCGTATGGAATCA
 CATGGTTGACACCTGCTGTGAATGGACCTCACCCCTGGCTTTCTAGGTAGGCTGCTG
 GAACATCGATACTGTCTTAACGTTGTCTGCAACGGCTTTGAACAGTGTCTTTAGAAA
 TTAAACCCTTATGTTAATAGTGACTTTTAGTTGCTACGATGTTGCTAACAGTCTGGTT
 TACCGTCTTGTTACCATAGACATTGACTGTAGCGCTGTAAGTTGAGTACCAATTGCTC
 GGAACACTGGTCAATGAGCGCTTGTCTTGCCAGCTACATTCCGCCAAGGCTACT
 TGATAAAATTATTGACCTTGGC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	5	661	R	219 aa

[SEQ ID NO:] 3858314-6 ORF translation from 5-661,
 direction R
 VIPYELFAGDGMLTRLLLKASDKAPWSDNGDAKNPALSPLGENVKTKQYFYQVALDGNV
 AGKEKQALIDQFRANGTQTYSATVNVYGNKDGKPDLNDNIVATKKVTIKINVKETSDTANG
 SLSPSNSGSGVTPMNHNHATGITDSMPADMTSSTNTMAGENMAASANKMSDTMMSEDKA
 MLPNTGETQTSMASIGFLGLALAGLLGGGLKNKKEEN*

Blastp and/or MPSearch Result:

Description:

Probable cell wall associated protease

Assembly ID: 3858368

Assembly Length: 2138bp

[SEQ ID NO:] 3858368 Strep Assembly -- Assembly
 id#3858368

CTTCCAGAACTTCTAAACCAGCCTCCATGATTACTGGGCCATTCCGTCTCCTAATTAGG
 AGCTACTATTTCTTGCCATAGCCTCTCCTTACACACTAGGCATATCGTGGTAAGAA
 ACACTGCGTCCCCTCACCTGCATTCTCTTTGAACAAAGGTATTAGCGTTATATAG

GCAATAGCAGAAGCCTCAACACATCAAAATCAAGCCCTGCTGCATTAAAGATGGTTCT
 GTATCTCTGTTCAACAGTGACCAAAACCCGATCCTGGCATCGATTCCATCTGTTACC
 GCATTGATAGTGTAGGACACCAAACGAACAGATTGGTTAAAGAACTTATCGATAGCGTTA
 AAGATTGCTCAACGGAACCTGCCCTGCGATTAAATCGACTTCTCACCATCCATA
 TTGGCTAGGCTAACGAGCGCTTCAATGTCATTATCTGCATGAGTTGAAGTTGAAATCA
 TCAAAGTGGAGCCTCTGGATTTCAACCATGGTCCAGCTACCAAAGCTCGAGTATCT
 GCATCTGTGATTTCTTACTTCTTATCGGCCAGTGCCTGAACCTAGCAAAGAAATGGTTG
 ATATCCTCTTCTGTAAGGCAATTCTCAGTTCTCAACAAAAGCATGGCGA
 CCAGATAATTTCAGCGGAATCTAACACCAACCAATTCAAGGTGTGATGATCTCATAA
 GTGAGAGGATTTAAGGACTCCATCTGGTGAATACCAAGATTGCTGGGAGAAGGTATTG
 CCACCAACGACGGCTTGTAGGAACATGGTTCTACTGGTAGTAATCTGGCGA
 GACGTATTGATGGTCTCATTTAGGACAATACTGGTTCTACTGGTAGTAATCTGGCGA
 ATATTGAGAGGCAATCGCAATCTTCAAAGCAGCATTCCAGCTCGCTCCCTAATACC
 ATTGATAGTCTCTCAACACGTCCTGCACCATTCTGACAGCAGCAAGGCTATTGCCAC
 TGCCATTCCGAGGTATCATGACAGTGAGGCGAATAGATGATCTGACGATCCGTCTGAC
 ATTCTCAATCAGGTATTGAAGATGGCACCACTTCTCTGGTAGGTAAATCCTATATT
 TTCTGAAAATTCTCAGTAAAGAATATTAGCTAATTGAAAGTTCATGAAAATTATTAA
 AATATTCATTTTAGAGGTTAAGTCCAACCTTTCTATCAATTCCAGTACTTCTTC
 ATCTGATAAAAGTATCATCAAGGGACACACTAATCCAGTAGCGCTTGCTCATATGGAAGGC
 TGGATAAAATCCCCTTGTGAAAGCAAATTAGCTACTTGGCATGCTGAGGTTGACTGC
 TTCCACTTGTCTCTGCCCTTCCAGCTATTCCAAGAGATTTCATCAAGACGGC
 ATACCACTTTGATTGCCCTCATGGCGCAATACAGCTGTATCAGGCGATTCCCACAG
 ATACTCCAATGGTTCCATACTTCTGAACTTGAGTCATGATACGCTTAGTCTGATG
 ACAGATAAAATCTTGACATCAAAACAAGCCTCCGAATCTGGTAAAGAATCTCCAGACA
 AGCCTCACGGACATTCGGACAAAATTCCCCTCATGCTTCCATATGAACGTGAGGATAA
 AGGTCACCAGTCTTGGTCAAAGACTGGAAAGTTCAACATTATCAGCAGTGATGGACAC
 AGTCATGACAAAGTCACCTTGCAAAATCTGGCAACTATATGTCCAGAATTCCCTATTT
 CTATAAAAACCATAATCATGAAGCCTTTCTGATTAAATTGATAGGATTAAAAATT
 TCAAACATAAGTGAAAACTGCTACCCAAAGGCTTAGCAGTTCTTCTATTTTAAAA
 AACAAACCTTAGTACCATGCAATTGTGTTACCCCCACCTGGTCAATAAAGGTTGACGGTT
 GTCAAGGTCAATCCCCCACCTGGTAGAATTCAATTTCACCTTACGCTACTCCAAAAT
 TCTGTGATAGTGAACAAACGTTTCTAAGGAGTCGCCAGACACACCAGCAGGTTAG
 GATACGAGTGACACCGGCTGACTGCCAGTCATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
9	1207	1578	R	124 aa

[SEQ ID NO: 3858368-9] 3858368-9 ORF translation from 1207-1578, direction R
VQDFICHQTKRIMTQVQEKYGNQLEYLWEKSPDTAVLRHEGNQKWYAVLMKISWNKLEKG
REGQVEAVNLKHDQVANLLSQKGIYPAFHMSKRYWISVSLDDTLSDEEVLELIEKSWNLT
SKK*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3858556

Assembly Length: 735bp

[SEQ ID NO: 3858556] 3858556 Strep Assembly -- Assembly
id#3858556
ACAGCTCACATCACTGTAGCTGTCAGAAAAATAAGGAGGTAAAATCGTGGGTCAAAAA
GTACATCCAATTGGTATGCGTGTGGCATCATCCGTGATTGGGATGCCAATGGTATGCT
GAAAAAGAATAACGCGGATTACCTTCATGAAGATCTGCAATCCGTAATTGTTCAAAAA
GAACTTGCTGACGCAGCAGTTCAACTATTGAAGTCGAACGCGCAGTAAACAAAGTTAAC
GTTTCACTTCACACTGCTAAACCAGGTATGGTTATCGGTAAAGGTGGTCTAACGTTGAT
GCACCTCGTGCAAAACTTAACAAATTGACTGGAAAACAAGTACACATCAACATCATCGAA
ATCAAACAACCTGATTGGATGCTCACCTGTAGGTGAAGGAATTGCTCGTCAATTGGAG
CAACGTGTTGCTTCCGTCGTGCACAAAAACAAGCAATCCAACGTGCAATGCGTGTGGA
GCTAAAGGAATCAAAACTCAAGTATCAGGTCGTTGAACGGTGCAGATATGCCCGTGCT
GAAGGCTACTCTGAAGGAACTGTTCCGCTTCACACACTTCGTGCAGATATCGATTACGCT
TGGGAAGAAGCAGATACTACATACGGTAAACTGGTGTAAAGTATGGATCTACCGTGGT
GAAGTCCTCCCAGCTCGTAAAACACTAAAGGAGGTAAATAACCAATGTTAGTACCTAAA
CGTGTAAACACCCGT

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
6	49	702	F	218 aa

[SEQ ID NO: 3858556-6] 3858556-6 ORF translation from 49-702, direction F
VGQKVHPIGMRVGIIRDWDALKWYAEKEYADYLHEDLAIRKFVQKELADAATVSTIEVERAV
NKVNVLHTAKPGMVGKGGANVDALRAKLNKLTGQVHINIIEIKQPDLDALVGEGLA
RQLEQRVAFRRAQKQAIQRAMRAGAKGIKTQVSGRLNGADIARAEGYSEGTVPVLHTLRAD
IDYAWEEADTTYGKLGKVWIYRGEVLPARKNTKGGK*

Blastp and/or MPSearch Result:

Description:

30S RIBOSOMAL PROTEIN S3 (BS2). - BACILLUS
STEAROTHERMOPHILUS.

Assembly ID: 3858562

Assembly Length: 1965bp

[SEQ ID NO: 3858562] 3858562 Strep Assembly -- Assembly
id#3858562
CTGTGTGATTCCATTATTGTCAAAATACTTTAGTTTAGCAATAACGACTTGCAGACA
AGACCAAGAGGGCAATCNANTTGGCAGAGCCATCAAGGCAGTAAACGATATCTGCGATAA
TCCAGACCATNTCCAACTCGATAAAATCCTCCTAACAAAGACCAGTGGCACAAAAACCACNC
GGTAGAGCCAGATAAAGCGAACCCAAAGAGGAACCTCAAAACAGCGTCTTCCGTAATAG
TTCCAACCTAGAATCGTTGAAAGCACAAGGAAGATGGTCAAGAAGGCAGGC
CCAAAGTGTAAAAGACTGTTGAGAAAGCTGACTGAGTCAAGGCAACCCATTCAAGTCA
CCACTCCAAACTCCAGTTACCAAGATGGTCAAACCAAGCTTAGAGTACAAATGATGAGGGTA
TCAATAAAGGTTCTGTCAATGGAAATCAAACCTTGCTCTACTGGTTCATTTGTCTTGGCA
GCTGCAGCTGCAATAGGAGCAGAACCCAGACCAGATTGTTGAAAACACACCACGCGCC
ACACCATTGAAATAGCCATCCGAACGCTAGCACCAGCAAATCCACCTACCGCAGCAAGG
GGACTAAAAGCTGAGGTAAGACTAAAGCGATTGTGCCAGGGATTTCCGATATTAAAG
AAAATAACTGTAAGAGTTCTAACAGATATAATGATGGCCATAAAAGGAACACAGTAGTT
GAAACCTTAGAAATAGACTTGAGTCCACCAAAGACTGCAATCGCTACAAAGACAGACAAG
ACGAGAGCTGTGATGGCTGGCGAAATCGCTGTATTTGGATAGATTCTGTAATCGAG
TTGACTTGGGTGAAGGTTCCGATTCCAAAGAGAGCAACCAATACTCCTGCTACTGCAAAC
AAAACAGCAAGTGGTCGCCACTTTCTCCATCCCTAGAAGGATATAATGCATGGACCT
CCCGCTACTGCACCATGGTCGTCTGGTGCAGTATTGATGGCCAAGAGTCCTCCGCA
TACTTGGTAGGCCATTCCAAAGAAAGCCGCCATCCACATCCAAAATAGAGCTCCCTGGTCCA
CCAACCTTGATAGCCGTCGCCAACTCCCTAACATGAATATTCCCTGTTCCCAACCAGTT
GAATGCCAAGGGCTGTTACACAAGCTGAAACTGGATAACATCACCATGTCCCTT

ATCCTGGATAAAATAAGCTGAAAGGCCTGGCAGACGCAAAACCTGCAAGAGTCCTAG
 CCGCATGGTTAGGTAAATCCCTGTCGACCAATAAATCAAGAGGGCGGTCCCCAAGCA
 AAAGCATCGATTGATTAAAGCAATTCTAACATTCCTCTCCTATCGTTCAACCCCCAAA
 AGAAAGAGCACATGCAAGATACTGTACTCTGGAATGCTTAGATAAAATGCTAAAAAGCGG
 TCTATCCTAGCTCTGCCTTACCTGAGAGTTGAGCAGTGCCTGCCTGCCCTTCG
 GTGCCTTACGGTCTCCAGAGTCCGTCATTACAGTCATGGAAAATCAAACGATTC
 CCCACTTCTATTAAACTTCATTGGTGTGGTATTAAATTGATTCTAATTACAAAAAA
 TGTTGGCTTGTCAATGTGTTATTAGTAAAAATTAGTTCAACAGTTTACTTTATAA
 AGTCCAGAATACTGCTATCCTTAAAGTACAATAGTCGCACTGCCTCCAGCATT
 TGTGGGCATAGCCGAAACTCTGACATGTTGTCTTGCAGTTATCTGGTAACCTCC
 TTCACGGGATGACTCCTGTTCCGATACCATGGGATGACATCAACTCGAAGCCCTATATT
 GTTAACCAAAGCTTGGTCAATGAAGGTATCTAGCCCATTCATGGCTCTTCATAGCGCT
 TGCCTCGAAGATTCACTGAGTCTAGCTGAGTCCTGCCAGAAGTTCG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	14	178	R	55 aa

[SEQ ID NO:] 3858562-6 ORF translation from 14-178,
 direction R
 VVFVLMVLLGGFIELXMVWIIADIVNALMALPXXIALLVLSQVIAETKKYFDK*

Blastp and/or MPSearch Result:

Description:

D-alanine permease (dagA) homolog - *Haemophilus influenzae*
 (strain Rd KW20)

Assembly ID: 3858656

Assembly Length: 1187bp

[SEQ ID NO:] 3858656 Strep Assembly -- Assembly
 id#3858656
 ACGTTGTCAATTAAATTATGAAACTAAGAGAAAAATTGTTAGGAAGCAGTAAATTGGT

GTCAGATAATGAAACAATAATGATAGAATCTGGATCGACCTGTGCTTACTTGCTGAGGA
 AATTGCAAGCAAAAAGAAATGTTACGATTGAAACAAATTGTTCTGGTGGTGGAGTTTCAGAAAGATT
 TGTGAGAGCTTATGATTGATGTCGTGTTATTGTTCTGGTGGTGGAGTTTCAGAAAGATT
 ACAGGTGACTGTAGGACCTTATTAAAAGAAATGATAACAGACTTTCATGTGTCAAGC
 TTTGTTGGGACAGATGGTTACGATAAAGAGATGGGCTTACCGGAAAAGATTAAATGCG
 CAGTGAGGTAGTTCAATATATTCAAGCAGTGTGCGATAAAGTCATTGTCCTAACTGACTC
 AAGTAAATTGATAAAAGAGGTACAGTAAGAAGATTGCTTAAAGTCAAGTCTATGAAGT
 AATAACAGACGAAAAACTTCTAAACAAAATATAGCTACATTAGAAAATGCTGGGATAAT
 GGTAAAGGTAGTTCGTAAGAGGTTAAGTGTATGAATCAAGATAGGAATAACTGCTTTC
 TAAAATTGCTTATCTGTATTATATTGAAAATTAAATCAGTCACAAATAGCAGCAAAATT
 AGGAATTATAGAACCTCTATTAGTAGAATGTTAACAGAACAGGAATGTAGGAATTGT
 TAAAATTGAAATAGAGAATTGATACCAATATGTTAAGTTGGAAAATTATGTAAAAGA
 AAAATACAGTTGGAAAGTTAGAAATTATTCAAATGAATTGATGATACTCCAACAAAT
 TTTATCTGAAAGAATTCTCAAGTTGCAGCAGGCGTCCTAGGAATCTAATTGATGATAA
 TATGAAAATTGGCTTCTGGGGAAAAGTTAAGTAATTAGTAGATTAAATTCACAG
 TAAAAGTGTCCGAAATGTTCACTTCTATCCTCTAGCAGGTGGCCTAGTCACACACGC
 TAAATACCATGTGAATACACTGATTATGAAATGTCTAGAAAATTTCATGGAGAGTGTAC
 ATTATGAATGCAACGATTGTGCAAGAAAATTGTTAGCAGATGGTATTGCAATC
 AAGATATTGAAAATTGAAAATAGTTGGAAAGATTAGATATAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	245	559	F	105 aa

[SEQ ID NO: 3858656-6] 3858656-6 ORF translation from 245-559, direction F
 VTVGPLLKEMIQTfhvcqafvgtgdgydkemgftgkdlmrsevvqyisavsdkvivltdss
 KFDKRGTVRRFALSQVYEVITDEKLSQNIATLENAGIMVKVVS*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3859118

Assembly Length: 843bp

[SEQ ID NO: 3859118] 3859118 Strep Assembly -- Assembly
id#3859118
AGCTATTGCAGGAACCAAGATNATGATTGGTACGTGGAGTTGGTATTATTNTACC
TCAAATCCTNGCAAATATGATTGGTTGACTACGATTCTGGTTAATCAATCAAATTAT
TACTTATGGGGTTATTGCGCGGTTGTTATCTTCTCTCCAGAGATTGGACTGGTTTGG
AACGTTGGGAAGAGCGACAGATTCTTCCAATGCCCTATTAGTGCTGAGGAACAGA
TGATTCGTGCCTTGTAAAGTCTGTCGAATACATGAGTCCTCGTAAATCGGGGCCTTGG
TTGCCTATTAGCGTGTACCGTACCTGCAGGAGTATATTGACAGGAATCCCCTTGA
TGCTAAGATTCTGCAGAACTTCTCATTAACATTATTCCAAACACTCCCCTACATGA
CGGTGCGGTGATTATCAAAGAAGAACGTATCGCTGTGACGTCTGCCTATCTGCCCTTGA
AAAAAACACAGGTATTCCAAGGAATTGGGACCAGACACCGGGCGCTATCGGTTTATC
AGAAGTCTCAGATGCCTGACTTTGCTATCAGAGGAAACGGGAGGAATTCGATAAC
CTATAATGGAAGGTTAACGACAACCTAACACTTGATGAATTGAAACAGAATTACGTTG
AAATCTTACTTCAAAGAGGAAGTGGGCCTTAGTTAAAGAACGAATGGCTAGGAG
GAATGGAAACATGAAAAAAATAGTTATATATCATATCCTCACTCCTTTTGCTTGT
GTCTTATTGCTATGCTACGGCGACGAATTCAAAACAGTACCAAGTGCTAGGCAGGTT
AAA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	314	661	F	116 aa

[SEQ ID NO: 3859118-6] 3859118-6 ORF translation from 314-661,
direction F
VYRTLQEYISTGIPLDAKISAELLINIFIPNTPLHDGAVIIKEERIAVTSAYLPLTKNTG
ISKEFGTRHRAAIGLSEVSDLTFVVSEETGGISITYNGRFKHNLTLDEFETELR*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3860084
 Assembly Length: 710bp

[SEQ ID NO:] 3860084 Strep Assembly -- Assembly
 id#3860084
 ATCGAATTAGTTGGTTGATTACCTCCAAGAAAAACTAGCCCTCTAGCCTTACTA
 GGAGCTGGTTGGTTACTAGTCTGTATTCGCCTTATCAGGTAAAACGTCAGATGCAG
 GACTAACATTGCTGATACGACACTAAAAAAGAAGTTGAGTCAGTTGTCTCAGCTTCTT
 TTTTGTACTACAGGATAATGGTGGTCCGTAGAGACTTAACTCTTCGAAAATCTCTTC
 AAACCACGTCAAGCGTCGCCTTACCGTACTCAAGTACAGCTGCGCTAGCTCCTAGTTT
 GCTCTTGATTCTCATTGAGTATTAACCTGGTCTGACTGGTCAAAGTGGAAAGCGGTCA
 TAGGCCCGCCAAGCGCGCGAGTTGGAGCATCTGGATCAAGAGCGCTGAGTCCCAGTAGA
 AGACTGGAAGTCTGGTAAAATTCTAGTTCAATCAAGAATCGATTATCCACTGTTCA
 GCCTGGCTAGAAAACCAAGAATAGAATTAAATTCGATCCCTGAAAGCGGACGTCGTCAG
 CGCTTGCCTGTTGCATGCTTGGTAGGCTTAAAGTCAGTAATCAAAGTATGAGCTC
 TTTTGATGGGTCTGTATCTGCATGGAAATGCCCTTTAATCTGGTGCCAGTCTTAC
 TTCTGGCAACTGTGTTTGATACTGTTAGTTATCAGCTTTAATTCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	294	473	R	60 aa

[SEQ ID NO:] 3860084-6 ORF translation from 294-473,
 direction R
 VDNRFLIELEKFYQTSSLLMGLSALDPDAPTRAARAYDRFHDPVKLILNENQRAN*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3860172
 Assembly Length: 1975bp

[SEQ ID NO:] 3860172 Strep Assembly -- Assembly
 id#3860172

CTTGATCTGACCGATGACACGTTGTGCAGTTAGCTTCAAGTTAAGTATTCTTGG
 CATCAGTCTGAGTCAGTTTGAAACGGGGATACCTGACAAGCGACTCAAGGTGGTCAAAA
 TATCAGACTCTGTCACCAAGTCTTATAGACAGGCACCCCTCTTGTGCGATTAGCT
 GGGCTGCCTGTTCCACTTGCCATCCATCAAGGCCTGTCAAGCTGGACTCAAGTCAGAAT
 CGTCTGCTTTACATGCTTGATTATTTGCACGTGTGCTGCCGCCTCATCCAAGAGAT
 CGATAGCAGAGTCTGGCAAGTGACGACTGGTAAATAACGATGAGCCATCTAACCGCTG
 TTTCAACCGCTTCATCTGTGATTGTACACGGTGATGTTCTCATAGTCGCCTCAAAC
 CTTGTAAAATAGTCATACTATCTGCCACACTGGTTCTCAATCGTCACTTAGCGAAC
 GACGAGAAAGTGCCGCATCTTTCGATATGTTTGATATTCTCCTGAGTGGTGGCAC
 CAACCGTTCTCAAAGTCCACGCGCAAGGCTGGTTCAAGATATTGGCCGATCCAGAG
 TCGAATCAATTCCGCTACCAAGAACCCATGATGGTGTGGAGTTCATCGATAAAAGAGGATGA
 CTTGGCCATCTCTTCAATATCCTGATGATATTATTATCGCGTCTTCAAAGTCACCAC
 GGAAGCGTGTCCCTGCAACGACATTCAAAATCAAGTTCTAACACCGCAGTAGCCA
 TTTCCGCAGGCACGTCAACACTGGCAATACGCTGGCAAGACCAAGCGCCAGAGCTGTT
 TCCCGACACCAGCATCCCCAACCAAGACAGGGTTGTTCTAGTCTTCCGGCTAAGATT
 GAATCATACTGTGAGATTCTCTTGTCCCCGACCGATGACTGGTTCTAACTGCCAGAACCG
 CTTGCTCTGCAAATCATGCGTATAGTCCTCAAGACCACACTAGGAGTCTGCGGCATGC
 CCATCATATTGCCATAGAATTTCGTTGTCAGCTACTGTACGATGGCGTTGGCGCAAAG
 CCTTGAGATCTCACGAGTCCAGCCTGCCGTTCTAAATTGACGAAGAGCAGCAA
 TCTTGACCTGATCTTCTTGTCTTCATAAGAAAAACCAGCCCTCTCCAAGATACTGAGTC
 CCAAGGCATTGCCATCATGCAAATCGCATAGAGGACGTGCTCTGCCCTAGCACCTAG
 CATGGACCACTGACACTACATACTCTGCTTCGTCAAAAGAACCTGCAAACGACGGGAGA
 ACGGCAATTCCGTAAAGGTTCATCCTGGCTATAGTCCGTTCACTCAGTTCCAAGGCA
 CCTCTTCTAAACGGTCCATCTCATACGGATAATCATTAAAGTTGCCCTGCTACACTAT
 AACTGTGATTAGACATGGCAATCAACAAGTGCCAAGACTCTAGATAACGAGGCTCCAAA
 TGTCAGCAACCAGTAGGCACCTTCGATACATTCAATGCTTTGAATAGTCATC
 TTACTTCCCTTTCTATCTACCTCTGTATGACCTGACGTAGCATGTTGCTCGAACAA
 TGGAGCTTCTCTCCTAAAACGCGATCCAAAGCTACTGATTCTAGCAAATTCACTCTC
 CTTGGTCATCAATTCTGCTCAACCAAAAGCTGGAGAATATCCTCATAAATTGAC
 TGACTCGCTACCAATCGAGTAAAGCAGCTCCCGAACATTCTATGATGACTAGAAA
 CAATCCGCTCTATACGAATGTAGCCTCCACCACCGCTTACTTCAACCAAGTAGC
 TACTTCCGTAAAGCGTGTCTTGATCACGTAGTTAATCTGACTAGGAACAAACCTG
 TATCTGCCAACTGACTCCGTTGCAACTCCACGATAACCAGATTGATCTAAATCGC

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
8	1724	1888	R	55 aa

[SEQ ID NO:] 3860172-8 ORF translation from 1724-1888,
direction R
VIKTRFTESRGYLVESKRGGGGYIRIGRIEFSSHHEMFRELLYSIGERVSHRNL*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3860242
Assembly Length: 1592bp

[SEQ ID NO:] 3860242 Strep Assembly -- Assembly
id#3860242
GCCCATTAGGGTAACCTTTGCAGCCTAACAGGCGATTGATTTCTGGCCAC
GAATCTGGATTATTATTTAACAGTAAGAGGAAATTATGACTTTAAATCAGGCTT
GTAGCCATTTAGGACGTCCAATGTTGGGAAGTCAACCTTTAAATCACGTATGGG
CAAAAGATTGCCATCATGAGTGACAAGGCGCAGACAACGCGAATAAAATCATGGGAATT
TACACGACTGATAAGGAGCAAATTGTCTTATCGACACACCAGGGATTCACAAACCTAAA
ACAGCTCTCGGAGATTCATGGTGGAGTCTGCCTACAGTACCCCTCGCGAAGTGGACACT
GTTCTTTCATGGTGCCTGCTGATGAAGCGCGTGGTAAGGGGACGATATGATTATCGAG
CGTCTCAAGGCTGCCAAGGTTCCGTGATTTGGTGGTGAATAAAATCGATAAGGTCCAT
CCAGACCAGCTTGTCTCAGATTGATGACTCCGTAATCAAATGGACTTTAACCGGAAA
TTGTTCCAATCTCAGCCCTCAGGGAAATAACGTGTCGCTAGTGGATATTGAGTG
AAAATCTGGATGAAGGTTCCAATATTCCCGTCTGATCAAATCACAGACCATCCAGAAC
GTTTCTTAGTTTCAGAAATGGTCGCGAGAAAGTCTTGCACCTAACCTCGTGAAGAGATT
CGCATTCTGTAGCAGTAGTTGACTCTATGAAACGAGACGAAGAGACAGACAAGGTT
ACATCCGTGCAACCACATGGTCGAGCGCGATAGCCAAAAGGGATTATCATCGGTAAAG
GTGGCGCTATGCTTAAGAAAATCGGTAGCATGGCCCGTGTGATATCGAACTCATGCTAG
GAGACAAGGTCTCCTAGAAACCTGGGTCAAGGTCAAGAAAAACTGGCGCGATAAAAAGC
TAGATTGGCTGACTTGGCTATAATGAAAGAGAAATACTAAGTAGAGGGTAGGCTCATGCC
TGCTTCTTGTGTTTACAGAAGGAGGACTTATGCCTGAATTACCTGAGGTTGAAACCGTT
GTCGTAGCTTAGAAAATTGATTATAGGAAAGAAGATTTCGAGTATAGAAATTGCTTACCC
CCAAGATGATTAAGACGGATTGGAAGAGTTCAAAGGGATTGCCTAGTCAGATTATCG
AGTCAATGGGACGTCGTGGAAAATATTGCTTTCTGCCTGACAGACAAAGGTCTGATT
CCCATTGCGGATGGAGGGCAAGTATTTTATTATCCAGACCAAGTGCCTGAACGCAAGC

ATGCCCATTTCTCCGGTTGAAGATGGGGCACGCTTGTATGAGGATGTACGCA
 AGTTTGGAACCATGGAACCTCTGGTGCCTGACCTTGTAGACGCCTACTTATTCTAAAA
 AATTAGGTCTGAACCAAGCGAACAGACTTGTACAGGTCTTCAAGCTGCCCTG
 CCAAGTCCAAAAGCCTATCAAATCCCCTAGACCAAGACCTGGTAGCTGGACTTG
 GCAATATCTATGTGGATGAGTTCTCTGGCGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	573	1001	F	143 aa

[SEQ ID NO:] 3860242-7 ORF translation from 573-1001,
 direction F
 VSRLVDILSENLDQYFQPSDQITDHPERFLVSEMVREKVLHLTREEIFHSVAVVDSM
 KRDEETDKVHIRATIMVERDSQKGIIGKGGAMLKKIGSMARRDIELMLGDKVFLETWVK
 VKKNWRDKKLDLADLGYNEREY*

Blastp and/or MPSearch Result:

Description:

GTP-BINDING PROTEIN ERA HOMOLOG. - STREPTOCOCCUS MUTANS.

Assembly ID: 3860282

Assembly Length: 1604bp

[SEQ ID NO:] 3860282 Strep Assembly -- Assembly
 id#3860282
 TCATCAAAAGCAGTTAACGAATTGTGAGCGTGTATTGAGAAATCATGAAAGTACGGAC
 CGATACATATAAAAGGATTAACTATGGAAGAAATTCTCTGTATTGGTTGTGGAGCAACC
 ATTACAGACGACAGATAAAGCTGGCTTGGTTACCCCCCAGTCGGCACTTGAAAAAGGT
 TTGGAGACTGGCGAAGTCTATTGCCAACGCTGTTCCGTCTCCGCCACTACAATGAATCA
 CAGATGTCCAGTTGACGAACGATGATTCCCTCAAGCTCTTGCACGAGGTGGGAGACAGTG
 ATGCTTAGTGGTCAATGTCATTGATATCTTGATTTAATGGATCTGTCATCCCAGGTT
 TACCACTTCTCGTCTCGGCAATGATGTCCTTGGTAGGAAATAAAAAGATATCCTTC
 CTAAGTCAGTTAAGTCTGGTAAGATTAGCCAGTGGCTCATGAAACGTGCCATGAAGAAG

GTCTTCGTCCAGTCGATGTGGTCCTAACCTCAGCACAAAATAAACATGCCATTAAGGAAG
 TCATTGACAAGATTGAACACTACCGTAAGGGCCGCGATGCTATGTGGTCGGTGTGACCA
 ACGTTGGAAAATCAACTCTAACATGCTATTATCCAAGAAATCACGGGTGATCAGAATG
 TCATCACTACTTCACGCTCCCAGGGACAACCTGGACAAAATAGAGATTCCGCTTGACG
 ACGGATCTTATATTTACGATACGCCGGAAATTATCCACCGTCACCAGATGGCTCACTACT
 TGACGGCCAAAACCTCAAGTATGTCAGTCCTAAAAAGGAAATCAAGCCTAACGACCTATC
 AGCTTAATCCTGAGCAAACCTATTTAGGTGGTTGGGACGCTTGACTTTAGCAG
 GAGAAAAGCAAGGATTACTGCTTCTTGATAATGAACCTCAAACCTCCATCGTAGCAAGC
 TTGAAGGAGCTAGTGCTTCTACGATAAGCACCTGGGAACTCTTCTGACACCACCAAATA
 GCAAGGAAAAGAAGATTCCCAAGGCTAGTCCAGCATGCTTACCATAAAGATAAGA
 CAGACCTAGTCATCTCAGGCCTAGGATGGATTGTCGTAACAGGCACAGCAAAGTCGCCG
 TCTGGCACCAGAAGGCCTCGCCGTCGTACACGAAAAGCAATTATTAAGCACAGAAAG
 GAAAGGGTTGTCTGAATTGGCGAGCAAGGCGAGCCCCATAGAGAATACTTCGCTGT
 GGTGTAAGTGGTACAAGTGATTGATCAACTGCGGAAAATTGAGACCTTAGGCTCAA
 TTTAGTCATGAAAGTCCGAAGGACTTTGCTGACGTCCGTACCACTCAGAAAAGTATA
 AAAAGAAAATCTTTAAAGAAATTATGTCATTAACATCAAACACGTGCCTCCTCAAC
 AGCCAGGCACACACCCCTCAAACCTATCCTAACATCGGGAAAATGGACTCAACGACCAA
 ATCAAACACAGCGTCCGTCAAGCTTGTGATGCCCGTTGAATTATCAAGGTTACTCCCC
 TTTACAAAACACAGATTGAAAACATCCGGACGAATGTAATTG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	288	1190	F	301 aa

[SEQ ID NO: 3860282-6] 3860282-6 ORF translation from 288-1190, direction F
 VGDSDALVVNVIDIFDFNGSVIPGLPRFVSGNDVLLVGNKKDILPKSVKSGKISQWLMKR
 AHEEGLRPDVVLTSQAQNKHAIKEVIDKIEHYRKGRDVYVGVTNVGKSTLINAIQEI
 GDQNVITTSRFPGTTLDKIEIPLDDGSYIYDTPGIIHRHQMAHYLTAKNLKYVSPKKEIK
 PKTYQLNPEQTLFLGGGLGRFDFIAGEKQGFTAFFDNELKLHRSKLEGASAFYDKHGTLL
 TPPNSKEKEDFPRLVQHVFTIKDKTDLVISGLGWIRVTGTAKVAVWAPEGVAVVTRKAI
 *

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3860296

Assembly Length: 2025bp

[SEQ ID NO:] 3860296 Strep Assembly -- Assembly
id#3860296
CCGTAATGGGTCGTAACCTGCCCTTAATATTGAATCACGTGGTTACACAATTGCTATCT
ACAACCGTAGTAAAGAAAAACGGAAGATGTGATTGCTTGCCTCCTGAAAAGAACTTTG
TACCAAGCTATGACGTTGAAAGTTGTAACACTCAATCGAAAAACCTCGTGTATCATGC
TGATGGTTCAAGCTGGACCTGGTACAGATGCTACTATCCAAGCCCTCTTCCACACCTTG
ACAAGGGTGATATCTTGATTGACGGTGGAAACTTCTACAAAGATACCATCCGTCGTA
ATGAAGAATTGGCAAACCTCAGGTATCAACTTTATCGGTACTGGAGTTCTGGTGGTGAAA
AAGGTGCCCTGAAGGTCCTCTATCATGCCTGGTGGACAAAAAGAGGCCTACGAATTGG
TTGCGGATGTTCTTGAAGAAATCTCAGCTAAAGCACCAGAAGATGGCAAGCCATGTGTGA
CTTACATCGGTCTGATGGAGCTGGTCACTATGTGAAAATGGTTACAATGGTATTGAGT
ACGGTGATATGCAATTGATCGCAGAAAGCTATGACTTGATGCAACACTTGCTAGGCCTT
CTGCAGAGGATATGGCTGAAATCTTACTGAGTGGAACAAGGGTAATTAGACAGCTACT
TGATCGAAATCACAGCTGATATCTTGAGCCGTAAAGACGATGAAGGCCAAGATGGACCAA
TCGTAGACTACATCCTGATGCTGCAGGTAACAAGGGAACTGGTAAATGGACGAGCCAAT
CATCTCTGACCTTGGTGTACCATTGTCACTGATTACTGAGTCAGTGTGACGCTACA
TTTCAACTTACAAAGAAGAACGTGTACATGCTAGCAAGGTGCTCCAAAACCAGCTGCCT
TCAACTTGAAGGAGACAAGGCTGAATTGATTGAAAAAAATCCGTCAAGCCCTTACTTCT
CAAAAATCATTACGCACAAGGATTGCTCAATTGCGTAGCCTCTAAAGAAAACA
ACTGGAACTTGCCATTGAGATATCGCATCTATCTGGCGTGTGGCTGTATCATCCGTT
CTCGTTCTTGCAAAAGATTACAGATGCTTACAACCGCGATGCAGATCTTGCCAACCTTC
TTTGGACGAGTACTTCTGGATGTTACTGCTAAGTACCAACAAGCAGTACGTGATATCG
TAGCTCTGCGGTTCAAGCAGGTGTGCCAGTGCCAACTTTCTCAGCAGCTATTACTTACT
TTGATAGCTACCGTTAGCTGACCTCCAGCTAACATTGATCCAAGCACACGTGACTACT
TTGGTGTGTCACACTTACCAACGTAAAGACAAAGAACGTTCAACTACTCTGGTATG
ACGAAAAATAAGTAGGTCAGCCATGGGAAACGGATTTTATTACTTGAGAAAGAACGAAA
TCTAGCTCATTGTTAAGTTGAACTCCAGAAAGAGCAGTATCGGTTGATCTGGTAGA
GGAGGGGCAAAAGCCCTCTCCATGGCTCTCAGACAGACTATGATTGATTTATTGAA
TGTTAATCTGGGAGATATGATGGCTCAGGATTTCAGAAAAATTGAGCCGAACAAACC
TGCCTCAGTCATCATGATTAGATCATTGGAAAGACTTGCAAGAAGAGCTGGAAGTTGT
TCAGCGTTGCGAGTTTCATACATCTATAAGCCAGTCCTTATCGAAAATCTGGTAGCGCG
TATTTGGCGATCTTCCGAGGTCGGGACTTCATTGATCAACACTGCAGTCTGATGAAAGT
TCCAAGGACCTACCGCAATCTAGGATAGATGTTGAACATCACACGGTTATCGTGGTGA
AGAGATGATTGCTCTGACACGCCGTGAGTATGACCTTGGCGACACTTATGGAAAGCAA
NGAAGTATTGACTCGTGAGCAATTGTTGAAAGTGTGTTGAAAGTATGAAAGTGCACCGA
GACAAATATCGTAGATGTCTATATCCGCTATCTACGGAGCAAGCT

ORF Predictions:

ORF #	Start	End	Direction	Length
8	1697	1843	R	49 aa

[SEQ ID NO:] 3860296-8 ORF translation from 1697-1843,
 direction R
 VMFNIYPKIAVGPWNFHQTAVLINESEDRRNTRYQIFDKDWLIDV*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3860406
 Assembly Length: 1578bp

[SEQ ID NO:] 3860406 Strep Assembly -- Assembly
 id#3860406
 CTACACCGGTTGGTAAAAATCGTATGCAAACCAAGGAGGCTGGAGTCAGATTGATGT
 TCAGTTGAAACGTCGAAATGACCTCTTGCCTAACTTGATTGAGACTGTAAAAGGTTATGC
 CAAATATGAAGGTTCTACCTTGAAAAGGTGGCAGAACTACGTAACCAAGTGGCGGCAGCG
 AATTCAACCAGCAGAAGCTATGAAAGCCAGTGATGCCCTCAATCGTCAGGTTCAAGGTATT
 TTTGCAGTTGCAGAAAGCTATCCAGATTGAAAGCTAGTGCTAACTTTGTTAAATTGCAA
 GAGGAGTTGACAAATACAGAAAATAAAATTCTTACTCTCGTCAACTCTATAACAGTGT
 GTCAGCAACTACAATGTAAAATTAGAAACTTTCCCGAGCAATATTATCGCTGGAATGTT
 GGATTAAAGCGGCAGATTCCCTCAAACACCTGAAGAGGAAAAGTCGGTTCTAAAGTT
 GATTTAGCGGTTAGGTGACTAAGATGTTGATCAAATTGCAAGCAATAAACGAAA
 AACCTGGATTTGTTGCTGGTATTTCTACTCTTAGCTCTGTTGTTATGCGGTTGG
 TTATCTCTTATAAGATCTGGACTTGGTGGTTATCGCACTGATTATCGGCTTTAT
 CTACGCTTGTCTATGATTTCAATCGACAGAGATTGTCATGTCCATGAATGGAGCGCG
 TGAGGTGGATGAGCAAACGGCACCAGACCTCTACCATGTAAGTGGAAAGATATGGCTCTGGT
 CGCTCAGATTCCATGCCCGTATTTCATCATTGATGATCCAGCCTAAATGCCTTGC
 GACAGGTTCTAATCCTAAAATGCGGCTGTTGCTGCGACTTCAGGTCTACTAGCTATCAT

GAATCGTGAAGAACTAGAAGCTGTTATGGGACATGAAGTCAGTCATATCGTAATTATGA
 TATCCGTATTCGACTATTGCAGTTGCCCTGCTAGTGCTATCACCAGTCAGTCAGAAGTGA
 GGCAGGTCGATGATGTTGGGGTGGAGCAGGTCGCAGACGAAGTGATGATGACCGAGA
 TGGAAATGGTCTTGAAATCATTATGCTAGTGGTTCCCTACTAGCTATTGTACTGGCACC
 TCTCGCTGCAACCTGGTTAGCTCGCTATTCTCGTCAGAGGGAATTCTGGCAGATGC
 ATCTAGTGTGAGCTGACTCGCAATCCCCAGGGAATGATGCTAGCAGTGCCTTATATCAATGC
 CAATAGCAAACCTATGAGTCGCCACGTCGATGATGCTAGCAGTGCCTTATATCAATGC
 TCCCAAGAAAGGTGGGGGTCCAAAACCTTTTATACCCACCCACCTATCTCAGAACG
 GATTGAACGTTAAAACAGATGTAAGGCTGGAAAAAGTCCTTAAAATCTGAAA
 AATGCATAATATCAGGTGTGAAAACCTGATATTATGCGTTTACTATGGGAAGATTACT
 TCTTTCTCCTAAAATTGTGTTTGCACCTATCTGCTATGTCATACTGGCAAATTGATAA
 ATCTTCTAAATTAACCTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	148	504	F	119 aa
7	497	1405	F	303 aa

[SEQ ID NO:] 3860406-6 ORF translation from 148-504,
 direction F
 VAELRNQVAAANSPAEAMKASDALNRQVSGIFAVAESYPDLKASANFVKLQEELTNENK
 ISYSRQLYNSVVSNYNVKLETFPSNIIAGMFGFKAADFLQTPEEEKSVPKVDFSGLD*

Blastp and/or MPSearch Result:

Description:
 unknown

[SEQ ID NO:] 3860406-7 ORF translation from 497-1405,
 direction F
 VTKMLFDQIASNKRKTWILLVFFLLLALVGYAVGYLFIRSGLGLVIALIIGFIYALSM
 IFQSTEIVMSMNGAREVDEQTAPDLYHVVEDMALVAQIPMPRIFIIDDPALNAFATGSNP
 QNAAVAATSGLLAIMNREELEAVMGHEVSHIRNYDIRISTIAVALASAITMLSSMAGRMM
 WWGGAGRRRSDDDRDGNGLEIIMLVVSLLAIVLAPLAATLVQLAISRQREFLADASSVEL
 TRNPQGMINALDKLDNSKPMMSRHVDDASSALYINAPKKGGVQKLFYTHPPISERIERLK
 QM*

Blastp and/or MPSearch Result:

Description:

HEAT SHOCK PROTEIN HTPX PRECURSOR. - ESCHERICHIA COLI.

Assembly ID: 3860416

Assembly Length: 1644bp

[SEQ ID NO: 3860416 Strep Assembly -- Assembly
id#3860416
TTTTTACCACTTCACCGGAGTTTTCTTCCTAACCTCCATCAGGATTAATCGCTGTAAA
GATACGTTCTTAACCAGTTTCCTTCTGTTCNACACGAGTTCACCTAGAAACAGT
GTTGAATCTTTCTCACTGCTTGAAGGCCAAATCTTTCAACAAAATTCGAGTT
GTGGGGAAAGATCTTCTTGTAAACAGCAGCAACTGTCTTCAGAAACTGGTTTTCCC
TTAGTCAACTGGATAACGGTATTCCTAACCTGTTCCACTTCTGAAACAGGAGCGAAC
AAGTACTGGAAGGCAATCTCTCCACTATCTACCACAGTTGAAGCTACTTGATTGTTTC
TTCAACTGAGACTTTGGCCGTGACCTTATAGTAATTGATAGTCTTGACGATTTC
AGCGAAATCAGCAAGTCTTTCCATCTACAAGAATCTCGATTGCGTGCTTCTGAGG
CAATTCACTGGTGCAAGGAAGGTATCTCAATCATCGAACACCGCTTTATGCTTT
ACGCTCCATACGCCATCTCATAGCTTGCTTGACAGCTTAAATGTTACGTTGATTTC
ATCACCAAGCTGCGATGCTTTATCCGCACGATAAGGCACAGCTTCCAATTTCGATT
GTTGAATGGATGGTCTGCGTCGTAGGCTGGTAGTTGAATAGTAGGTTGGCACTTC
CTCTGGACCGACATAGCGTTCTAAAACGAGTTAGTTGGTCATCCGTACCGACTATCTGC
AAAGAAGTGAAGTTGGCTTGCACAGTCCGTTCTACAATCTACCATTTCACGGAA
GATCACACCCGCTGATACTTCTGGATTAGAAGATGGTGGAGACCAGTTGTCCAACG
ACGATTTCGATGATCTCCGTATTGAGATAGTCAACGCGGTATGAGAGTTTGTC
AATATCATTGGTTGCTGAAGCAAAGGCCTGGTTACTGTTTCATCATAGTTAGGGTTATC
TGAAAGAGCTTCGCTAGTTGTCGACTCGTACAGTGACCTCAGCAACAAAGATCACT
ACCAAGGACATGGCCTCGAACGGTAAATTGACCTGCTTGTCAGATTTCTGCTGGAAC
TTCTCCCATTCAACTGACAAATCTTGTTCGCTAGCCGTCTTACCTGTGAAGTAAAC
TGGAACCTTAGTCGGCAATTCAAGTGCTTGACCTACTTGTAGCAAGCGAGCTGTTAAC
CGCAGCAACTGGTTATGAGAAAGTAAGTCTTACCTAGTGAAGTGCAGACGGTATTC
TCCTAAGATGTCGCCATTTCAGCTTCCGATGACACGAACGGCTCACCTCACGAAC
GCTTGGAACGACGGTAGCGAGACCATTGTTGCTAACACTGGCTGTGACTGCCGGAACTT
TCCCACATCTACAGACTCAAGGTAGTATCTGTCAGATCAGGTTGAAGTTGCTAAGTCTTAA
CCGTCAACTTGGATTCTGTTGCTTGCTGGCTGCCGCAACTTGTTCGCAAAGATT
GTACCTCTGTGATAACGTTCTAATTGTTGTCTGCTCTCACCAGGCGAACAG
CATAGTTCAACTTATCAAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	72	281	R	70 aa

[SEQ ID NO:] 3860416-6 ORF translation from 72-281,
 direction R
 VENKLRNTGIQLTKGKTSFWRKTVAAVTRKIFPTTRNFVEKDLAFKTVEKKDSTLFLGET
 RXEQEGKTG*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3860712
 Assembly Length: 1087bp

[SEQ ID NO:] 3860712 Strep Assembly -- Assembly
 id#3860712
 ATCGAATTGCAAGTATGGCCATTGTCTTCATGTTAGTTCTTTAAGACTGTAAATC
 AAGGAATCCCTTACTATTCACTAGCGTAACTGATTCTACAGGATCCATTACTAATCTAC
 GCGCCGGGAAGTAGGCTGAGACATAACCAAGTAATAGAGCGAAAATAGAGTTCTAAAA
 CAGATAAAAGATTTAATTCAAAAACCTAGTGTGGATGGTAAAGTGACTTACAATCG
 CATTGCCAAACTTCCCACCCCTGTGCAACCAAAATGCCAGCAGCAAGGCGATGCCTA
 CAATCCAGATAGCCTCGTAAATAAAATCCTTGACATCACGATTCTGATAACCAACTG
 CTTCATGACACCTATTCCTTGGAACGTTGCATGATATTGATGTAATAATGATACCAA
 TCATAACCGCTGCTACCACAATAGCTTGTGATGAAAGCACAATCAATAATCCCTGAATAA
 CACGAATAAAGGTAATCACAATATCAAGAACTCTCTGTTAAGAAAGCACAGTATACTTCT
 TATTTCTGTAATTCTCTGTTACTACTTTGCTGTGATGGATCTTGAGTTCCAAGA
 TAAAATAAGATAACAGCTTCGTAATCCAGCCTCTTCAAAATCGTTCCATTGATGAG
 ACAGCATGAAACTGTTGCTGTCCTCCATGTCATCTTCAATTGATTACACGTACAATCT
 TCGTTGAAATTGAGCAATCTTACTAGTTGGCAGCACTTCTACAATGCTGACTGAGA
 CTGATTGCCAATAAGATCATTAGCTGTCATAATTCTGCTGTTCAATTCCAATT

TTAGTAAACTGCTTCCAATCGTTAATCCCTGTTCAATTGTATCAGTATAGAGGGATCCAG
 CCAACACTTGTCCGTCATTATTACAAACAGAGATACTTGTATCATCATAAAGACTCA
 CTACTTGAGCATAAGAACATCGTTGACTCAAATCCATTCTGCCATCTTCTTGC
 CCATCTATAGTAATATTGACATGTTCATCCAAAAGGACTCTCCAAATATTAATAGAT
 CGAGCCT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	74	499	R	142 aa

[SEQ ID NO:] 3860712-6 ORF translation from 74-499,
 direction R
 VITFIRVIQQLLIVLSSQAIIVVAAMIGIIIIYINIMQRSKEIGVMKAVGYQNRDVKGIFI
 YEAIWIVGIALLLAFLVAQGVGSLANAIVSHFYPHSITKVFELNLLSVLGTLVFALLLGYY
 SAYFPARKISKMDPVESLRYE*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3860728
 Assembly Length: 1283bp

[SEQ ID NO:] 3860728 Strep Assembly -- Assembly
 id#3860728
 ATCGAATTGAAAAATACAGCATGCCCTTGTCCAATTGGTACTTGAAAAAGGAGAAGAAG
 ACCGTATCTTTCAGACTTGACTCAAATCAAGCAAGTTGTTGAAAAAACAGGTCTGCCCT
 CTTTTTAAACAAAGTGGCAGTAGACGAGTCGGATAAGGAAAAACGAATTGCTTTTC
 CAAGATTCTGTGTCGCCCTTATTACAAAACATTATCCAGGTTCTGCCCTACAATCACAGA
 GCAAATCTTTTATGATGTGCTTGTAGATTGCTTGAACCGACTTGAAAAAGAAACAAAT
 CGATTGAAAGTGACGATTACGTCTGCTCATCCTCTAACTGATGAACAGAAGACTCGTTG
 CTCCCTTGATTGAGAAAAAAATGTCTCTGAAAGTAAGGAGTGTAAAAGAACAAATCGAT
 GAAAGTCTCATTGGTGGTTTGTCAATTGCCAATCACAGACAATTGATGTGAGTATT

AAACAACAACCTAAAGTTAAAGAAAATTGAAATAGAAAGTGGTGTCTTTGGCAA
 TTAACGCACAAGAAATCAGCGCTTAATTAAGCAACAAATTGAAAATTCAAACCCAATT
 TTGATGTGACTGAAACAGGTGTTGTAACCTATATCGGGGACGGTATCGCGCGTGCATG
 GCCTGAAAATGTCATGAGTGGAGAGTTATCGAATTGAAAACGGCTCTTATGGTATGG
 CTCAAAACCTGGAGTCAACAGACGTTGGTATTATCATCCTAGGTGACTTACAGATATCC
 GTGAAGGCAGATAACATCCGCCGTACAGGGAAAATCATGGAAGTCCCTGTAGGTGAAAGTC
 TGATTGGTCGTGTTGATCCGCTTGGTCGTCCAGTTGACGGTCTGGAGAAAATCCACA
 CTGATAAAACTCGTCAGTAGAACGCACAGCTCCTGGTATTGCAACGTAAGTCTGTT
 CAGAACCATGCAAACCTGGTTGAAAGCTATTGACGCCCTGTACCGATTGGTCGTGGTC
 AACGTGAGTTGATTATCGGTGACCGTCAGACAGGGAAAACAACCATTGCGATTGATAACAA
 TCTTGAACCAAAAAGATCAAGATATGATCTGTATCTACGTCGCGATTGGACAAAAAGAAT
 CAACAGTTCGTACGCAAGTAGAAACACTTCGTCAGTACGGTGCCTGGACTACACAATCG
 TTGTGACAGCCTCTGCTTCACAACCATTGCTCTTCCTAGCTCCTTATGCTGGGG
 TTGCTATGGCGGAAGAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	259	519	F	87 aa

[SEQ ID NO:] 3860728-6 ORF translation from 259-519,
 direction F
 VLVDCLNRLEKETNRFEVTITSAHPLTD EQKTRLLPLIEKKMSLKVR SVKEQIDESLIGG
 FVIFANHKTIDVSIKQQLKVVKENLK*

Blastp and/or MPSearch Result:

Description:

ATP SYNTHASE DELTA CHAIN (EC 3.6.1.34). - ENTEROCOCCUS
 FAECALIS (STREPTOCOCCUS FAECALIS).

Assembly ID: 3860794

Assembly Length: 1402bp

[SEQ ID NO: 3860794 Strep Assembly -- Assembly
 id#3860794
 CTAATCAATCCAAAAGGAGCAACCAAATAACTGGTCCACCATTCCCAATGAGCATCTGCA
 AAAAGTTTCAACCCATAGCTGGCAATGCAATATTAAGAATGTCTTATTCTTAAAC
 AATCTCTCCTCCTGATGAAAAGAAACTCAGTTGGTTCCCAACCGAGTTACTCCCTCT
 ATCTTAAAGTCCTAAATAAGCCTCAACCGCTACTTGCATGTCAGCAGCTGCCACTGTTGT
 CTTGTGACGAACAGGAGCTGTCTCAAGCCCATCAACTGCTGGTGGCACTGCAACGCCTGA
 GATTTCATGTAATTGAGCAAAGCTTCAAAGTCTGTTAACCTGCTTTCCAGTTACAGC
 TTCTACTGCAACTACTGGAACTTGTAGGGACTAGCTGTTGAAGCAATCACTGTCTTAGT
 CGCATCATCAGTAACCGCTGGTATTCTATAAAACTGCTGAGGCAACGCCGTATGTGG
 ATCCTCAATATAAGAATCTAACTCATAAACACGCTTGATTCTGCCGCTGTTCTTCCTC
 AGTCGCATATTCAAGCTCAAAGAGCTCCAGAATCTCTACATCAAATCAGTCAGTTCAT
 TTGTCCTGTGTATTCAAGGTATTCAAGTTCAGCTGAGGCTTAACCGCATCATTCCCCAA
 AAGATGGAAAATCAAACGCTCAAAGTTGAAGATAACCAAGATATCCATAGATGGGCTGGT
 TGTTACCTTAAACTCACGTTCTGCTAAACACGTGCTTGAAGAAAGTCTGTCAAAAC
 ATTGTTATCATTGAAGCACAGATCAATTACCAACTGGGAGACCGATTTGTTGGCATA
 AAAGGCAGCCAAGATATTCCAAAAGTTCTGTTGGTACTGTGAAGTTAATCTTAC
 CAGCCACGATCTCACCAGTCTGACCAACTGAGCCATAGGCCATAAACATTAAACA
 ATCTGTGGCACCCAAACGACCGATATTCATAGAGTTAGCAGATGAAAATGCAACCT
 TGTTGGCGCTAATCTTCACGAAGAGCCACGTCGTTAACATGTGCTTCACGTTGGTT
 GCGCATCGTCAAAGTTACCATCTATAGCGATAACATGAGTATTGTCACCATTATGAGTGG
 TCATTTGCAACTCTGTACCTTGCTGACACCACCCCTTGGATAAAAGACGATAATCTCAG
 TACCAAGGCACATCCGCAAACCCGCCATAGCAGCTTCCCCGTGTCACCAGATGTCGCTG
 TCAAGATAACAATCTGTTCTCAAACCATGTTTTAGCAGCAGTCGTATAAGTATG
 GCAAAATAGACNAGGCCATATCCTAAAGGCAATNGTGAACCATGGAAAAGTTCCAAAT
 TGTATTGCCCATCTAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	184	915	R	244 aa

[SEQ ID NO: 3860794-6 ORF translation from 184-915,
 direction R
 VRSWLVIRLTSQYQQETFGNILAAYAKQIGLPVGKLICASNDNNVLTDFFKTRVYDKKR
 EFKVTTSPSMDILVSSNLERLIFHLLGNDAVKTAELMNTLNTQGQYELTDFDVEILELFA
 AEYATEEETAAEIKRVYELDSYIEDPHTAVASAVYRKYQAVTDDATKTVIASTASPYKFP
 VVAVEAVTGKAGLTDFAEALAQHIEISGVAVPPAVDGETAPVRHKTTVAAADMQVAVEAY
 LGL*

Blastp and/or MPSearch Result:

Description:

Probable threonine synthase

Assembly ID: 3860830

Assembly Length: 989bp

[SEQ ID NO:] 3860830 Strep Assembly -- Assembly
id#3860830
CTCTCGTCACATGGAAGAAGTTGGATTCAAATCCTCAATCTGGTCCAGAGCCAGAAT
TCTTCCTATTAAAGTTGGATGAAAATGGGGACCCAACACTGAAAGTGAATGACAAGGGTG
GCTAATTGGATTGGCACCTTACTGACCTTGCGGACAACACACGTCGTGAGATTGTGAA
TGTCTGACCAAAATGGGATTGAAGTAGAACCGAGTCACCACGAGGTTGCGGTTGGACA
GCATGAGATTGACTTTAAGTACGATGAAGTTCTCCGTGCTTGTGATAAGATCAAATCT
TTAAACTTGGTTAAAACCATTGCTCGCAAACACGGACTTACGCAACATTATGGCGA
AGCCAAAATTGGTATTGCTGGATCAGGTATGCACTGTAATATGCTCTTGTGATGCAG
AAGGAAATAACGCCTCTTGATCCAAATGATCCAAAAGGAATGCAGTTGTCAGAACAG
CTTACCATTCCTAGGCCTTGTGATCAAGCATGCTTACAACACTATGCCATCATGAACC
CAACAGTTAACTCATACAAACGTTGGTTCCAGGTTATGAAGCGCCTGTTACATTGCTT
GGGCTGGTCGTAACCGTTGCCACTTGTGCGATCAGCGTACCTGCTTCACGTGGTATGGG
AACTCGTCTTGAGTTGCGTTCACTGGATCCAATGGCGAACCTTACGTTGCTATGGCTGT
TCTTTGGAAGTTGGTTGTATGGTATTGAAAATAAAATCGAAGCACCAGCTCCTATCGA
AGAAAATATCTACATCATGACAGCAGAACAGCGCAAGGAAGCTGGTATTACAGACCTTCC
ATCAACTCTTCACAACGTTTGAAAGCTTGACAGAACAGATGAAGTGGTTAAAGCTGCTCT
CGGAGATCACATCTACACTAGCTCCTGAAAGCCAAACGAATCGAATGGCAAGTTATGC
AACCTTCGTTTCACAATGGGAAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	176	286	F	37 aa

[SEQ ID NO:] 3860830-6 ORF translation from 176-286,
direction F

VNVLT KMGFEVEASHHEVAVGQHEIDFKYDEVLPCL*

Blastp and/or MPSearch Result:

Description:

Glutamine Synthtase SAGLNAR NCBI gi: 468507NCBI gi: 47374 -
Staphylococcus aureus.

Assembly ID: 3860984

Assembly Length: 817bp

[SEQ ID NO:] 3860984 Strep Assembly -- Assembly
id#3860984
ATCGAATTTATCCGTAAAGACCATTCAAGCACTTGGCAAGTAATGGGTGTGATTGATTCGT
CTAGATGCCTTGCTTATGCAGTGAACGAAATTGGATACTAATGATTCTTGCTGGAACC
AGATATTGGGATTATTGGACAAAGTTGAGATATCGCTGCTGAGTATGGGACAGAGCT
TTTACCTGAGATTCACTGAACACTATTCACTGATTCACTGAGTTAAAGATCAGACCATGATTACTA
TGTTTATGATTTGCTCTCCAATGGTGACACTTATACTCTTACAGTTCCAGAACAGA
GCGTTGGCTAAGTGGTTAAAGATGAGCCGATGAAGCAATTACGACGCTAGATAACCA
TGATGGGATTGGAGTAGTGTCAAGGATATCCTGACCGATGAGGAGATTGACTATGC
TTCAAATGAACTCTATAAGGTTGGAGCCAATGTCAAACGTAAGTACTCTAGTGCCGAGTA
TAACAACCTAGATATCTTACCCAAATCAATTCAACCTAACCTATTCAAGCGCTTGGAGAT
GATGATGTCAAGTATTTCTCGCTCGTCTAATTCAAGCTTGTCCCCAGGTATTCCTCAG
GTTTACTATGTGGGTCTATTAGCAGGCAAGAATGACTTGAAATTATTAGAAGAAACTAAA
GAAGGTCGAAATATTAATCGTCATTACTATAGCAACGAGGAAATAGCAAAAGAAGTGCAA
CGACCTGTTGTGAAGGCCCTCTCAATCTATTCTTCCGTAACCGTTCAGAAGCCTTT
GATCTAGAAGGGACTACTGAGATAGAGACACCAACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	113	520	F	136 aa

[SEQ ID NO:] 3860984-6 ORF translation from 113-520,
direction F

VEPDIWDLKDVRDIAAEYGETLLPEIHEHYSIQFKIADHDYYVYDFALPMVTLYTLYSS
 RTERLAKWLKMSPMKQFTTLDTHDGIVVVDVKDILTDEEIDYASNELYKVGANVKRKYSS
 AEYNNLDILPKINST*

Blastp and/or MPSearch Result:

Description:

sucrose phosphorylase (EC 2.4.1.7) - *Streptococcus mutans*

Assembly ID: 3861088

Assembly Length: 556bp

[SEQ ID NO:] 3861088 Strep Assembly -- Assembly
 id#3861088
 ATCGAATTGCTCTAATAACAAGTTTTGGTCAAAGACCCGTCTAGTGGGAAGCATC
 CCCATTCCAGATGGAGTTTCACGATCACATAATCAACGTGTTAAGGTCAACCTG
 ACGTCCACCTGCATAAGAAATAGCACTTGAGGTCTTGCCTCCACATTCAGTAAAGTGTGTC
 TTGAGATGACCTTACGGAAAGCAAGATACTTGAGGTCTTGCCTCCACATTCAGTAAAGTGTGTC
 TTTTGATATTGTGAGGCTGAACCATAATATCCTCTGAACGTCCACCATCGACTCAAT
 CGTTTCCCTGGACTTCAATGTGTCCTGCAAAGAGGGAACCATGATGATGCTAGC
 ACCGAAGCGGATAGACTTAGCAATATCACCGTGAGTACGAATTCCCTCCATCAGCGATAAT
 CGGTTACCGCAGCCTGGCACACCAGCGTAAGAGCAGCCAATGCCAACCACCTGTTA
 CCAAAACCACTTAACTTGGTACAAACCTTACCAAGGACGGATTCCGACCTTAGTA
 CCATCCGCACTAGCAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	46	474	R	143 aa

[SEQ ID NO:] 3861088-6 ORF translation from 46-474,
 direction R

VVGSWLLRWCAKAARKPIIADGGIRTHGDIAKSIRFGASMIMIGSLFAGHIESPGETIE
 VDGGQFRGYYGSASQYQKGAYKNVGGKRILLPAKGHLQDTLTEMEQDLQSAISYAGGRQV
 ADLKHDYVIVKNSIWNGDASH*

Blastp and/or MPSearch Result:

Description:

inosine-5'-monophosphate dehydrogenase (guaB) homolog -
Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3861138

Assembly Length: 528bp

[SEQ ID NO:] 3861138 Strep Assembly -- Assembly
id#3861138
AAAAAGCCAGAGGAGTGTGAGGAAGTGGAAAATCGAAAATTGTGAAGGATATCTTATT
TATCTCAAGTGTCTCAGCCGGCAAGTCAGGAGGACCTTATCTTGCAGAGATTGCAGG
ATACACTCTAGCAAATCGTGATACCTGTGTTGGCTAGCTGCCAATATGATTGGGTGC
AGAAGCGCGTGATTATCTTAATCTTGGCTTAGTTCCCGTGGTCATGTTAACCCAGTGC
TTCTGTCCTTGAGGATCTTATGAGGCAGAAGAAGGCTGTTGTCCTTGGTAGGTGTGA
GATCAACTAACGCTTATGAAACCATAAGGCTTGCTATCGTGACAGCAAGTGGCAGGAAC
AGACCATTACCTTGACAGGCTTCCAGCTCAGATTGCCAGCATGAGCTGGATCACTTGG
AAGGACGAATCATTAGGAAGGAAAGCAAATGAAACGAATAGTCTTGAACTTATTTTA
TCGCAACGACCTGGGTATATCTTTACCGCCCCCTAACCTGACCAGC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	42	437	F	132 aa

[SEQ ID NO:] 3861138-6 ORF translation from 42-437,
direction F
VKDILFLSQVSQPASQEDLYLARDLQDTLLANRDTCVGLAANMIGVQKRVIIIFNLGLVPV
VMFNPVLLSFEGSYEAEVGCLSLVGVRSKRYETIRLAYRDSKWQEQTITLTGFPAQICQ
HELDHLEGRII*

Blastp and/or MPSearch Result:

Description:

fms protein homolog - *Thermus aquaticus* (fragment)

Assembly ID: 3861256
 Assembly Length: 638bp

[SEQ ID NO:] 3861256 Strep Assembly -- Assembly
 id#3861256
 CTTAGGTCACTTTAAAATTCAAATTCCGCAAGAACATCTGCCACTGGTGACCAATT
 TGCTCCTCTGAATCAAATGATGGCAACCGTCTGATAGTCCATCTAAAATGCTACCAGG
 AATAGCAAAGATATCGCGTCTTCCATTGCTCGCTCACAGGTAATGAGACTACCTGA
 ACCGATCTTAGCCTCTGCTACAATCACACCACGACAAAGTCCAGCAATGATGCGATTACG
 GGCAGGAAAATCGAAATTTCAGAGGTTGTCGCCAGATCCATATTCACTTAGAGGCCAGAT
 GGTCAATTGCCGATGTAGTCTTGCAAGCGTTGGCTTAGGATAAAACACATCCAGTC
 CTGTTCCAATCACTGCAATGGTTTCCGCCATTCTGAAAAGCTGCCATATGAGCTGCTG
 TGTCAATGCCCTTGGCCAGACCACTGACAATAACCAGTTCAAGCAGCAATGCCACGAACGCTACCT
 TCCGGGAATTCAAGGAAGGTCAAGATTCCCTGTTAAAATAAAAATACAGGCGCATIC
 ATATTATTCACTCCAAATCCCCAAGGGATAACAAGTC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	13	207	R	65 aa
7	236	529	R	98 aa

[SEQ ID NO:] 3861256-6 ORF translation from 13-207,
 direction R
 VIVAEAKMRSGSLITCERAMEEGRDIFAIPGSILDGLSDGCHHLIQEGAKLVTSGQDVLA
 EFEF*

Blastp and/or MPSearch Result:

Description:

SMF PROTEIN. - ESCHERICHIA COLI.

[SEQ ID NO: 3861256-7] 3861256-7 ORF translation from 236-529, direction R
VGIRACSKQGAKSVEKVIQGLENELVIVSGLAKGIDTAAHMAAFQNGGKTIAVIGTGLDV
FYPKANKRQLQDYIGNDHHLALSEYGSGEQPLKFRFSCP*

Blastp and/or MPSearch Result:

Description:

SMF PROTEIN (FRAGMENT). - BACILLUS SUBTILIS.

Assembly ID: 3861262

Assembly Length: 1727bp

[SEQ ID NO: 3861262] 3861262 Strep Assembly -- Assembly
id#3861262
NCAAAAAATGTAGTGATTACGGGAGCAACTTCAGGAATCGGAAGCGATTGCGCGTGCTT
ATCTGGAGCAGGGTGAGGATGTCGTTCTAACAGGACGACGGATAGACAGATTAGAAATCC
TTCAAGTCGGAGTTGCAGTAAGCTTCCAAATCAAACCGTCTGGACTTTCCACTAGAT
GTGACGGATATGGTCATGGTGAAGACTGTTGCTCTGATATTCTAGAAACGATAGGGAGG
ATTGATATCTTGGTCAACAACGCCGGACTGGCTCTGGCTGGCTCCCTATCAAGACTAT
GAGGAGTTGGATATGTTGACCATGTTGGATACCAATGTTAAAGGTCTGATGGCGGTTACT
CGCTGTTCTTGCAGCAATGGTAAAGTCATCAAGGTACAGTTATCAATATGGGTCA
ACCGCAGGAATCTACGCCCTATGCTGGTGCCTGTTACTCAGCTACCAAGGCTGCGGTT
AAGACCTTTCGGATGGACTGCGAATTGATACCATCGAACGGATATCAAGGTGACAAC
CATTAGCCTGGATTGTCGAAACAGATTCTCAACTGTCGTTCATGGTATAAAGA
GCGGGCTGCGTCCGTTACCAAGGAATAGAACGCTTGCAAGCTCAGGATATTGAGACAC
AGTAGTCTATGTGACCAAGTCAGCCTGCCGTGTTCACTGAGATTACAGATATGACCA
CAATCAACAGGCACAGGTTCATGATTCATAAAAATAAGAAATTCCCTCGAAAAGTTA
CAAATTCTGTAACTTTGATTCCCTACGAATAGATAAGTAGGAGGAAGAAAATATGT
ATAATAAAGTTATCATGATTGGCGTTAACGTCTACACCAGAATTGACAAAACCAACA
ATGACAAGTCGGTAGCGCGAGCAACTATCGCTGTGAACCGTCGTTACAAAGACCAAAACG
GTGAACGTGAAGCTGATTGTTCAATATGGTCCTATGGGCCAGAACTAGCCAGAAA
CTTGCGCAAGCTACGCAACCAAGGTAGTCTCATTCCGTTGATGGAGAATTGCGTACCC
GTCGCTTGAGAAAATGGCCAATGAACGTGACCGAAGTACTTGTACAGGATTCC
AACTCTGGAAAGTCGTGCTAACGTGCCATGCGTGAAAATAATGCAGGCCAAGATTGG

CAGATTAGTCTTGGAAAGAAGAAGAATTGCCATTTAATACTCTCGAAAATCTCTCAA
 ACCACGTTAGCTTATCCACAAACATCAAAGCAATGCTTGAGCAGCCTGCGGCTAGCTTC
 CTAGTTGCTTTGATTTTATTGAGTGTAGTTACTTGATAGCTCGACCAAGTCTG
 AGCTTGTGTTCAAGTGAGTTAGGACTGTTCTCAAGAACCAATTTCGGCTTGCCCA
 GGCAGAGTCATTAACACGTGCAGCAGTGAATACCAACGCCTGTTGACGGATAAATGG
 CAAGAGGTCTTGTAGATAGCGAAAAGTTGATCGTGCCTGCATTGGCTACAGATGATAC
 TGTGACAAACTTGTCTGAAGGGCAGAAACGCCACGTGTATCAGACAAGTCAAGGGCAGC
 AGATAGCCAGTCAAGCAAGTTTCACTGTACCAAGGGATAGAGAAGTTGTAGACTGGAGA
 GAAAATCCAGATAGCATCCGCAACGAGAACTGCTTCACGAGCAGCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	181	594	F	138 aa

[SEQ ID NO: 3861262-6] 3861262-6 ORF translation from 181-594,
 direction F
 VTDMVMVKTVCSDILETIGRIDILVNNAGLALGLAPYQDYEELDMLTMLDTNVKGLMAVT
 RCFLPAMVKVNQGHVINMGSTAGIYAYAGAAVYSATKAAVKTFSDGLRIRYHRNGYQGDN
 HSAWDCRNRFLNCSFSW*

Blastp and/or MPSearch Result:

Description:

HYPOTHETICAL OXIDOREDUCTASE IN DCP 3' REGION (FRAGMENT). -
 ESCHERICHIA COLI. (BLAST)

Assembly ID: 3864150

Assembly Length: 3808bp

[SEQ ID NO: 3864150] 3864150 Strep Assembly -- Assembly
 id#3864150

AACTGGAACAAATATGGTTTGTCAAAACACCAATACCGTAAGGTTGACCGTGAAACAG
 GTGTTGTCACGAACGAAATTGTTGGTTGACAGCTGATGAAGAAGATGAATATACTGTAG
 CTCAGGCTAACTCTCGTCTGAATGAAGATGGAACCTTGCTGACAAAGATTGTCATGGGAC

GTCACCAAGGGTCAACCAAGAGTATCCAGCTAATATTGTTGACTACATGGACGTTCAC
CAAAACAGGTAGTGGCGTTCGACAGCATGTATTCTTCTTGGAAAACGATGACTCCA
ACCGTGCCCTCATGGGAGCCAATATGCAACGTCAGGCTGTGCCATTGATTAATCCTCAGG
CACCTTACGTTGGTACTGGTATGGAATACCAAGCAGCCCACGATTCTGGTGCGGCTGTGA
TTGCTCAGTATGATGGTAAAGTTACTTACGCAGATGCTGACAAGGTTAGAAGTCGTCGTG
AAGATGGTTCATGGATGTTACCACATCCAAAAATTCCGTCGTTCAAACTCAGGTACTG
CTTACAACCAACGCACTCTCGTAAAAGTTGGTATGTCGTTGAAAAAGGCGATTCATCG
CTGACGGACCTCTATGAAAATGGAGAAATGGCGCTTGGACAAAACCAATCGTTGCCT
ACATGACTTGGGAAGGTTACAACCTCGAGGATGCCGTTATCATGAGCGAACGCTTGGTGA
AGGACGATGTCTACACATCTGTCACCTGAAGAATACGAATCAGAAACGCGCAGACAA
AGCTTGGGCCTGAAGAAATCACTCGCAAATTCCAAACGTTGGTGAAGATGCCCTCAAAG
ACCTTGACGAAATGGGGATTATCCGTATTGGTGTGAGGTTAAAGAAGGTGATATTCTG
TAGGTAAAGTAACACCTAACGGGTGAGAAAGATCTTCAGCTGGAAAGAACGTCCTTGAC
GCTATCTTGGAGACAAGTCTCGTGAAGTGCCTGATACTTCTTCGTGTACACACGGT
GCCGATGGTGTGTCGTGATGTTAACATGTTGGTTCGTGTTACATCGCTAAAAACGTAAGGATTAAGGTCGGA
GATAAAATGCCGGACGTACGGAAACAAAGGGTTGTCTCTCGTATCGTCTGTAGAA
GACATGCCTTACCTCCAGACGGAACCTCAGTCGACATCATGTTGAACCCACTGGGTG
CCATCACGTATGAATATCGGTAGGTTATGGAGCCTCACCTTGGTATGGCAGCTCGTACT
CTTGGTATTACATTGCGACACCAGTCTTGATGGAGCAAGTCCTGAAGATCTTGGTCA
ACTGTTAAAGAACGAGGTATGGATAGCGATGCCAAGACAATCCTTACGATGGACGTACA
GGTGAACCATTGATAACCGTGTCTGGAGTCATGTACATGATCAAACCTCCACCAC
ATGGTTGACGATAAAATTGACCGCGTTCAGTCGGACCTTATTCAACTGTTACCCAAACAA
CCACTCGGAGGTAAAGCTCAGTTGGTGGACAACGTTCGGTGAGATGGAGGTTGGCT
CTTGAAAGCCTACGGTGCCTCAAATGTCCTCAAGAAATCTGACTTACAAGTCTGACGAT
ATCAACGGACGTTGAAAGCCTATGAAGCTATTACAAAGGCAAACCAATTCCAAAACCA
GGTGTCCAGAATCCTCGAGTTCTGTCAAAGAATTGCAATCTTGGTCTGACATG
CGTGCCTAGACGAAGATGACCAAGAAGTGGAACTTCGCGACTGGATGAAGGAATGGAC
GAAGATGTCATCCACGTAGATGACCTGAAAAAGCCCGCAAAAGCAGCCAAAGAGGCT
AAAGCAGCCTTGAAGCTGAAGAAGCTGAGAAAGCAACAAAAGCGGAAGCAACAGAAGAA
GCTGCTGAACAAGAATAAGCAGTTCACTTAGAATAGAAAGGGAAAGAAATAGTGGTTGATG
TAAATGTTTAAAGTATGCAAATCACCTAGCTCTCCAAAGTAAAGTCCGTTCATGGT
CTTATGGAGAAGTCAAAAACCTGAAACAATCAATTACCGTACCTGAAACCAGAACGTG
AAGGACTCTTGATGAAGTGATCTTGGCCTACAAAAGACTGGGAATGTGCTGTGGTA
AGTACAAACGCATTGTTACAGAGGAATTGTTGTGACCGCTGTGGGTTGAAGTAACGC
GTACGAAAGTTCGTCGTGAGCGTATGGGACATATCGAATTGAAAGTCCTGTATCTCACA
TCTGGTACTTCAAGGGATTCCAAGCGTATGGCCTGACCCCTGATATGAGCCCTCGTG
CCCTCGAGGAAGTTACTTGTGCGCTTATGTGGTATTGATCCTAAGGATACACCAC
TTGAGCACAAGTCTATCATGACAGAGCGCAATACCGAGAGCGCTTGTGAAATATGGTT
ATGGTTCATTTGTTGCTAAGATGGGTGCGGAAGCCATCCAAGACCTTTGAAGCAAGTAG
ATCTTGAAGAAATTGCTGAACCTAAAGAAGAATTGAAAAGTCTACTGGACAAAAC
GTGTCAAAGCCATCCGTCGTTGGATGTTGGATGCCTTTACAAGTCTGGAAACAAAC
CTGAATGGATGATTCTAACATCCTCCGGTTATCCCACCAGATCTCGTCCAATGTAGC

AGGAATTCGATGGTGGCCGTTTGCCTCATCTGACTTGAATGACCTTACCGCCGTGTT
 ATCAACCGTAACAACCGTTGGCTCGTTGCTTGAAGTAAATGCACCCAGGTATCATCGTT
 CAAAATGAGAAGCGTATGCTTCAAGAACGAGCAGTTGACGCTTGATTGACAATGGTCGTCGT
 GGTGTCCTAACACAGGACCAAGGTAGCCGTCCATTGAAATCATTGAGCCACATGCTTAAA
 GGTAAACAAGGACGCTTCCGTCAAAACTTGCTCGTAAACGTGTTGACTTCTCAGGACGT
 TCCGTTATCGCCGTTGGTCCAACCTTAAGATGTACCAATGTGGTGTGCCACGTGAAATG
 GCGATTGAACTCTTAAACCATTGTCATGCGTAAACGTTGCTTGCCTGATATCGTCAA
 AACGTCAAAGCAGCTAAACGCTTGGTGAACGCGGAGATGAGCGTATCTGGATATCCTT
 GAAGAAGTGATTAAAGAACACCCAGTGCTTGAACCGCGCACCGACCCTCACCGTTG
 GGTATCCAAGCCTTCGAGCCAGTCTGATTGATGGTAAGGCTTTCGCTTGACCCACTT
 GTCTGTGAAGCCTACAATGCTGACTTGACGGGACAAATGCCATCCACGTACCAACTT
 TCAGAAGAACGACAAGCAGAAGCTCGTATCCTCATGCTAGCTGAGCACATCTGAAC
 CCGAAAGATGGAAACCGGTAGTTACTCCATCTCAGGACATGGTTTGGTAACTACTAC
 TTGACCATGGAAGAACGCTGGTCGGAAGGTGAAGGAATGGCTTCAAAGACCGTGACAAA
 GCGGTTATGGCTTACCGCAATGGTTATGTTCACCTCCACTCACGTGTTGGTATCGCAACA
 GACAGCCTCAACAAGCCTGGACAGAACAGACATAAGGTCTGCTTACAACAGTT
 GGTAAAATTCTCTTCAACGATATCATGCCAGAGGGCTACCATACTTGCAAGAACCAAAC
 AATGCCAACTTGACAGAACGCTGGTCCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	922	1998	F	359 aa
8	2031	2759	F	243 aa

[SEQ ID NO: 3864150-7] 3864150-7 ORF translation from 922-1998, direction F
 VRKIFQLEERLLHAI~~FGDKSREVRDTS~~LRVPHGADGVVRDV~~KIFTRVNGDELQSGVNMLV~~
 RVYIAQKRKIKVGDKMAGR~~HGNKGVVS~~RIVPVEDMPYLPDGT~~PVDIMLNPLGVPSRMNIG~~
 QVMEPHLGMAARTLGI~~HIATPVFDGASPEDLW~~STVKEAGMDSDAKTILYDGRT~~GEFDNR~~
 VSVGV~~VMYMIKLHHMVDDKLH~~ARS~~VGPySTV~~TQQPLGGKAQ~~FGGQRF~~GE~~MEVWA~~EAYGAS
 NVLQEILTYKS~~DDINGRL~~KAYE~~AITKGKPIPKPGVPES~~FRVL~~KELQSLGLDMRVL~~DEDD
 Q~~EVELRDLDEGMDEDV~~I~~HVDDLE~~KAREKAAQEA~~FEAEEAEK~~ATKAEATE~~AAE~~QE*

Blastp and/or MPSearch Result:

Description:

DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6)
 (TRANSCRIPTASE BETA CHAIN). - BACILLUS SUBTILIS.

[SEQ ID NO:] 3864150-8 ORF translation from 2031-2759,
direction F
VVDVNRFKSMQITLASPSKVRWSYGEVKKPETINYRTLPPEREGLFDEVIFGPTKDWECA
ACGKYKRIRYRGIVCDRCGVEVTRTKVRERMGHIELKAPVSHIWFKGIPSRMGLTLDM
SPRALEEVIVYFAAYVVIDPKDTPLEHKSIMTEREYRERLREYGYGSFVAKMGAEEAQDLL
KQVDLEKEIAELKEELKTATGQKRVKAIRRLDVLDAFYKSGNKPEWMILNILPVI
PPDLR
PM*

Blastp and/or MPSearch Result:

Description:

DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6)
(TRANSCRIPTASE BETA' CHAIN) (FRAGMENT). - BACILLUS
SUBTILIS.

Assembly ID: 3864190

Assembly Length: 2753bp

[SEQ ID NO:] 3864190 Strep Assembly -- Assembly
id#3864190
ACCCGCTTCAGAACTAACAGATTGCGGATGTATTGTAATGGCAATCTATCTCTAG
AAGTCAGTGTAGTCCCTGCCTCAGAAAGTCCTAAAGAGCGAAGTGAGGGCTATCGTA
GTCAGGGTTACCAAGTACTGTGGTTGCTGGGTCAAAAAGTGTGGCTAAGGAGCGTTGA
CTCGTCTACAGCAAGGTTTCTTATTCAGTCAAAACATGGGCTTTATGTTGGAAAT
TAGACAAGGAAAAACAAGTTTAAGACTCAAATACCTGATTACCAAGGATCTCCGCGGT
AACTCCATTATCAAATCAAGGAATTTCTATGGTCAAGGTAGTTATTGAAATATTGC
GTCTTCCCTATAAGAGACAAAAATATCTCATTTCAGTTCTGAGGACAAGGACATCT
GTCGCTATATCCGGCAACAACTTTATTATCAAATCTCTTGGATGAAAGAACAGCAG
AAGCCTATCAAAGGGAGAAAATACCTGACTTATGGACTGAAAGAATGGTATCCACAAA
TTCGACCAATAGTGGCAAATTTCAGATTGAACAAGACTTGACTAGCTATTATCAGC
ACTTTTACCTATTACCAAAAAATCCTCAAATGATTGGCAAAGCTTATCCACCA
CCTTTTATCAGCAATATTCTGAAAAATATGGTAAAGATGAAAGGATGGAGGAATCTAA
TGGTATTACAAAGAAATGAAATAATGAAAAGATACATGGGATCTATCAACGATCTACC
CAACTGACCAGGCTTGGGAAGAAGCCTTAAAGATTTAACAGAACAAATTGGAGACAGTAG
CCCAGTATGAAGGCCATCTTGGATAGTGGATAACCTACTAGAAATCACTGAATT
CTCTTGAATGGAACGCCAGATGGAGAAGCTTACGTTATGCTCATATGAAGAACAG

AGGATACACGTGAAGCTAAGTATCAAGAGTACTATGCCAAGGCCATGACACTCTACAGCC
 AGTAGACCAAGCCTTTCATTCTATGATCCTGAATTATGGAGATTAGCAGAAAGCAGT
 ATGCTGACTTTTCTAGAACCTAACAAAGCTGCAGGTTATCAACACTATTTGACAAGC
 TCTTGCAAGGCAAGGATCACGTTCTTCACAACGTGAAGAAGAATTGATTGGCTGGAGC
 TGGAGAAATCTTGGTCAGCAAGTGAAACCTTCGCTATCTGGACAATGCAGGATATTGT
 GTTCCCTTATGTCCTAGACGATGGTAAAGAAGTTCAGCTATCTCATGGGACTTACAC
 ACGTTGATGGAGTCTAAAAACGTGAGGTTGCCGTGGCCTATCAAGCTTTATGC
 GACTTACGAACAATTCAAACACACCTATGCCAAAACCTTGCAAACCAATGTTAAGGTGCA
 AAATTGATGCTAAAGTCGTAACACTACAAGAGTGCTCGTCATGCAGCTCGCAGCGAAT
 TTTGTTCCAGAAAGTGTATGACAATTGGTAGCAGCAGTCGCAAGCATTGCCACTC
 TTACATCGCTATCTGAGCTTCGTTCAAAATCTGGGATTCAGATCTCAAGATGTAC
 GATGTCTACACACCGCTTCATCTGTTGAATACAATTTCACCAAGAACGCTTGAAA
 AAAGCAGAAGATGCTTGGCAGTCTGGTGAAGGATTACTTGAGCCGTGCAAACGTGCC
 TTCAGCGAGCGTTGGATTGATGTTACGAAAATCAAGGCAAGCGTTAGGTGCCTACTCT
 GGTGGTTCTTACGATACCAATGCCTTATGCTTCAACTGGCAGGACAATCTGGACAAT
 CTCTTACTCTGTTCATGAAAACAGGTACAGTATGCATTCAAGCTATACTCGTAAACT
 CAGCCTTATGTTACGGAGATTACTCTATCTTTGGCTGAGATTGGCTCAACTACCAAT
 GAAAATATCTTGACGGAGAAATTATTGGAAGAAGTGGAAAGACGACGCAACACGCTTGCT
 ATTCTCAATAACTCCTAGATGGTTCCGTGGAACAGTTCCGCCAAACTCAATTGCT
 GAGTTGAACACGCCATTACCAAGCAGATCAAAATGGGAGGTCTGACAAGCGATTTC
 CTAATAAAACTCTACGCAGACTTGAACCAAGAGTATTATGGTTGAGTAAGGAAGACAAT
 CCTGAAATCCAATACCGAGTGGCTCGCATTCCACACTCTACTATAACTACTATGTATAT
 CAATATTCAACTGGCTTGCAGGCCCTCAGCCTGGCTGAAAAAATTGTCCATGGTAGT
 CAAGAAGACCGTGACCGCTATATCGACTACCTCAAGGCAGGTAAGTCGGACTATCCACTT
 AATGTCATGAGAAAAGCTGGTGTGATATGGAGAAGGAAGACTACCTCAACGATGCCTT
 GCAGTCTTGAACGCCGTTAAATGAGTTGAAGCCCTTGGTGAAGGAAATTAGGATTGGCA
 TAAAATGGTTGAATCGTATAGTAAGAATGCTAACATAACATCGTCGTCCTGTCGCAA
 AGAAGAAATTGTAGACTTGATGCGTCAGCGTCAAAGCAGGTACAGGTTCTGAAAGA
 ATTGGAAGACTTGCCTGCAAGGAAATTCTATTATCCCCATGAAACGTTGCTTA
 TTCCGTTTCTATGAAACCAGCAGCTAAAATATTCTGGAAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
8	1259	1534	F	92 aa

[SEQ ID NO:] 3864190-8 ORF translation from 1259-1534,
 direction F
 VFPYVLDDDGKEVQLSHGYTRLMESKKREVRRGAYQALYATYEQFQHTYAKTLQTNVKV
 QNSMLKFVTRVLVMQLSQRILFQKVFMFIW*

Blastp and/or MPSearch Result:

Description:

ligoendopeptidase F - *Lactococcus lactis*

Assembly ID: 3864204

Assembly Length: 2140bp

[SEQ ID NO:] 3864204 Strep Assembly -- Assembly
id#3864204
CCAGTTTGGTTCTGCATGTTGTTAGGCAGGACGAGCGAGACGTTGGAAGTCTTCTTG
ATAAGCCAAGAGGCCAGATAACGGTCTTCTTATCCACTCAAGACGGATGTAGAGTTG
GTCGCCCTCTTAGGCCAGAGTCTTGAGCACAGGGAGAATATCGAGTGACAACAAACGA
TTTCCTTGTCAAGGAAGGCCTGTATCCACAAAGACACCCAAGTCCTACGAACCTCTGTGA
CACGTCCCCAACCAAATTGGTCCTGAGTGGCAGTCACCTCTAAGGTGTCAGGCAGGAGTT
TTTGCTTCATATCCGTGTATGCAAAACCTTGACCGTATCCCCTACTGTATGTTGCCCT
CTTCCTTAGCAAGAGCATAGGTTGACCATCCTTGCACAAAGTAAAACGGTCATTT
CATCGATGATCAGTCCAACGATAAAACTGCAAGATTGTATTCATATTCCTCTTCG
AATAAAACTCAGCCAGCAATGCCAAGTGGATTTCTGTTATTAGACTTCCAAAAG
TTCTTCTCTTGTAGCAGTCATGTCGATGTGTTAACAGCATCGTCTGTTACTTT
TTGAATATCTTTCAAGAGTCTCAATTGCTTCAGTGATTCTTGTCTTGTCTTCTTG
TTCTTAGCTCGTCATAGCATCGCAGGGATATTGCGGACAGCCACTTAGCATTTC
GCCGACCTCTTCACCTCTTAGCAAGGTACGACGAGTTCTCTGTAAGAGCTGGGAT
AACCAAGCGAATCACAGAACCGTCATTAGCCGGTGTGATACCAAGATCAGAACGTTCAA
GGCACGTTGATGTCTTCAATGAAGACTTGTCAAATGGTGTACCAACAAACACGCGC
TTCTGGAATCGTAATTGAAGCGATTGGTTAAGAGGGAGTTGACTCCATAGTATTCTAC
ATGTACACGGTCAAGCAAGCTTGCATTGGCACGACCAGCACGGATACCAAAATTCAAC
AGCAAGTGATTGGTGAGACTGGGTCAATTCTCTCTTAGCTTTCAATAATTACGTTAGC
CATATTCTTCTTATTCCCTTTCTCGATATTATTGAAACTGTTGTTCCGATATTTC
CCAAATACGACACGTTGATGTTGCCTGATTGGTCATGTTGAAGACAACCAAGTCATG
TCGTTGTCATTGAGAGGGTTGAGGCTGTTGAGTCCATGATACGAAGACCTTGTTGATA
ACATCACGGTGGGTCAATTCTCAAACCTAACGGCTGTCTGTCCTTAGGATCGGCA
TTGTCACACACCATCGACGCCATTAGCCATGAGGATGGCATCTGCTTCGATTCAGCT
GCACGAAGGGCCGCTGTTGATCTGTCGAGAAGTATGGTGAACCAATTCCAGCACCAAAG
ATAACGATACGGCCTTTCAAGGTGACGAAGGGCACGTCCACGGACATAAGGCTCTGCC
ACTTGTTGCATAGCAATAGCTGTTGACACGCGTATCAACCCAACTTGTTGCAATGAA
TCTGCCATCACAAGAGCATTACACAGTCCAAAGCATTCCAGTGTAACTGCCTGAACA

CGGTCCATACCTGCTTCTGCTGCAGGTTCTCCACGCCAGAGATTCCTCCACCAATAACA
 AGGGCAATTCGATACTAACGCTATGAACCTCTGAATCTCTTGCATTGTTGAAC
 GTTGGATATCAATCCCTACGCCACGTTCACCGGCAAGGGCTTCACCTGATAACTGATT
 AAAATACGTTATACTGGGATTCGCCATTTCACTCTCCTTCTTCATCCTACCTATTT
 TATCACACAATTCTAAGATTTATAGTATCATGAACAATTCTTCAAAAAAATTAGACAG
 TCAAAAATTCTCTAAGTCGGCAAGGGCACGCTCTGCAATTTCATAACGAGCCTCT
 TATCACGGATACGCTCGCCTTCCAACCTCCTGATGATGCCAAATTGACATTGAGGTT
 GGAAATGTTGCTGTCGGCATGGTAATGTAATGAGCTAACGTTCCAATCGCTGTCGTCT
 CGGGGAAAATAACCTCGCTTCTTCCTTGAAGAGACGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
8	1092	1835	R	248 aa

[SEQ ID NO:] 3864204-8 ORF translation from 1092-1835,
 direction R
 VKMANPKYKRILIKLSGEALAGERGVGIDIQTVQTIAKEIIQEVHSLGIEIALVIGGGNLW
 RGEPAAEAGMDRVQADYTGMLGTVMNALVMADSLQQVGVDTRVQTAIAMQQVAEPYVRGR
 ALRHLEKGRIVIFGAGIGSPYFSTDAAALRAAEIEADAIILMAKNGVDGVYNADPKKD
 AVKFEELTHRVDINKGLRIMDSTASTLSMDNDIDLVVFNMNQSGNIKRVVFGENIGTV
 NNIEKE*

Blastp and/or MPSearch Result:

Description:

URIDYLATE KINASE (EC 2.7.4.-) (UK) (URIDINE MONOPHOSPHATE
 KINASE) (UMP KINASE) (SMBA PROTEIN). - ESCHERICHIA COLI.

Assembly ID: 3864212

Assembly Length: 2545bp

[SEQ ID NO:] 3864212 Strep Assembly -- Assembly
 id#3864212
 CTCGCAGTTCTCCATAGCTAATTGCGCCAAACGTCCCTGCCAAGGTTGAGTCTGTCCCC

CAGAAATCCCTAGAACAAAGGTTTTAGGAAGGGATGTTTTCAGATATCTTTTAAGG
AAAATCAAATAGAACGGACGGATTCTCCGGTGGGCATCAATCACTGGTTGACAACC
CAGCTCTGGATAATCGGTTCTGGCAAACACTCATTGCTCTCCCTTCACCAAGGGCT
TCCTTGCGCATCTTATCAATCAAAGTCCATCTTATCTTGCCATACGTCACGCGCAAATC
CACTGGATAGTGCCTGGGATTGAGCACACGCTTAAACTCATCCCACAACCTGTCAAATTC
CTTACGGGCATAATCTGAATGTCAGTCAAACTAGGCAAGTTGAAACTAATATTCTTC
TTTGAAGATATCCACCAAGAGAGGAACGGCATCAAATTACGAACCGTCTTGTGATGTA
TGTATAGGTCGGATGGAACATCTGATTCTGTCATGTCGCTAATATCCACACCATCATA
AGTGATGTAGTCACCTCTGACTTGCTTTCACGACTGGTAATGCGCCACACCTGCTT
CTTACCTGGCGTCGACACTTTCCGCATTATTAGACAGCTTAATCGTATTGCGCATCTG
GCCGTTTCATCTCGATTGCAACAATCTGTAACCGCCCCAAGAGGCCGGCTGGTCATA
GGCTGTAATCAGCTTGGTACCCACACCCAGACATCAATCTTGGCTTTGCATCTTGAG
GTAAAGGATGGTATTTCATCTAGATCATTAGAACATAATCTTAGCCTCTGGAAATCC
AGCCTCGTCCAGTTGCTGACGGACTTCTTAGAAATGTAGGCAATATCCCCAGAGTCAT
CCGCACACCCATAAAGTTAATCTGATCACCCAGCTCACGCGCACCTGAATGGCAGCTGG
TACACCGATGCGAAGGGTATCATAGGTATCCACAAGAAAGACACAATTGATTTGTGGGT
CGCAGCGTAAGCCTTGAAGCCTCATAGTCATTGCCATAAACCTGTACCAAGGCATGGC
ATGGGTTCCAAAACAGGAATGTCAAAGAGCTTACCGCACGCACGTTGCTGGTCCATT
GGGCCACCAATCACCGCTGCCGTGTTCCAGATGGCCGCATCCATTCTTGAGCCCCGA
CGTGTCCAAAACCTCCATCAAGGGTCATCTCGATAACCAAACGAATACGAGTGCTTGT
CGCCACCAAGGTCTGGTAGTTGACGATGTTCAAAGAGCCGTTGACCAACTGACATTG
GGTAGAGGTCCCTCACCTGACAATCGGTCATTAGCAAAACCAAATCCCTCTTG
GGCAGAACGAACGGTCAACTCCAACCTGAAATTGCGAAGGTAATCCAAGAACGCCCATG
ATAACCAAGCGACTCCAAATAGGCTATATCACTATCTGAAAAACGCAAGTCTCAAGATA
GTTCACAAATTCTTCCAAACCTGCAAAACCGCATAGCCGTTCTAAAAGGCTGTTGGCG
GAAATACACCTCAAAGACCGCCTCTTATTGTAATCCCTGATCAAAGTAAACCTGCAT
CATGTTGATCTGGTACAAGTCCGTGTGCAATGTCAAACATCTGATCTGGATACATACTTT
CCTACTCCTTAGCTAGAAACCCATGAAAATTTCAGAACACTTTCATGTATTCCAATAAA
TTAGTACTATTATACACATTAGCTGGATTGAGAAAAGAGTAACAGCTATTCTCAC
TCTCCAATTCATCCATATCTGTTCAAATTCTTGAGGCCATTGCCATAGCTTTAA
GACCAAGATTGCCATAAAGACCCACGGAAAGGTAATGACATAAGTAATGACCCAAGCAG
ACAGGTATTAAAATCAAAGGATTGTGCTGATAAAATTCTATGTTGAATTGATAATTCT
GCAACATCAAAGAGCCGTAATAGCAAGGTTAGGAAAAACACCCAAATCGTAAAT
GAAAACGACTATAGTAGGTCACTCCAGATAACGGGCACGATTGAAAAAGTAAATGTCC
CTATGATGATAACGATTAGCAGCATATTAGAATTAAAAGGCTGGTCTAATACTGAAA
TGATATAAGATAGGAGCGACAAAGCAATGCAGATATAGAAACTTTAGAGGCCGCTTAT
TGAACAGTTGTTCTCTTCTGCTAGTAATTGATAATAAAAATCTATTCTCATCT
TCTCCTCCAAAATAGTTGGCTAGGGTTTCCCTAAACATCTGCAAATAGACTGGCAG
AGCGAGAGACTGGGATTGTATTCTCCGCCTCTATCAAACCAATAGTCTGGCGTGTCA
CCGACAGCCTCTGCCAGTTGACCTTGTGTTAAATCACGCTCTACCCGAGCTAATTAAAT
TTTAAATTCTTAGGCCACCTTCGTCCTCCTTATAGTTAAACTCATCTACGCTTAAAAA
ATCCAAAACCAACACAAGCTATCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	256	1155	R	300 aa

[SEQ ID NO: 3864212-6] ORF translation from 256-1155, direction R

VIGGANGTSNVRAGKLFIDIPVLGTHAHALVQVYGNDEAFKAYAATHKSNCVFLVDTYDT
 LRIGVPAAIQVARELGDQINFMGVRIDSGDIAYISKVQRQLDEAGFPEAKIYASNDLDE
 NTLNLKMQKAKIDVWGVGTLITAYDQPALGAVYKIVAIEDENGQMRNTIKLSNNAEKV
 STPGKKQVWRITSREKGKSEGDYITYDGVDISDMTEIKMFHPTYTYIKKTVRNFDAVPLL
 VDIFKEGILVYNLPSLTDIQDYARKEFDKLWDEFKRVLNPQHYPVSDLARDVWQDKMDFD*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3864214
 Assembly Length: 3655bp

[SEQ ID NO: 3864214] Strep Assembly -- Assembly
 id#3864214

ACTTGATTAACAAATTAAACCTGCTAACTGCATCCAACGAATTCTGGATCTTAGCTTG
 GTTGCTCCTCCCTGCCATGGCCATGTCCTGGTTACCACCAACGTCCATCGATGATTA
 GTGCTAATTCTTGACAAGGTTCCCTGCATGAAGGTCTTGTCTGCTTGCTACAAGGA
 CATTGACTTTGTCACCGATAGCGGCAACTAGGACAAGAAGATCAGAGTAGTCTTTGTT
 TCCAGTTATCTGCAAAAGTACGAAGGGCACCGGCATCGGATAACAGACACTGACTAGCAA
 TGTAACGATGACCGTTGACTTCCTTAACATCTTGAAGATATGCCCTCGGGCTGCAGCTG
 CGGCTTTCTTCAACTCAGCATTCTTGAAGGTTGACGAAGTTGTTCTGAAGTC
 CTTCTACCTGTGAGGTACTTCCTGACTGAGGTGCTTCAAGGTTGCTGCGACAGCTT
 TAAGAGCATCCTTTGACGATAGGCTCAAGGCTTCTTACCAAGTCAGTCACTGCCAAGA
 TACGGCGAGTTCCCTGAACCGATTCTCTTGAACAATTGAAAGAGACCAATCTCAG
 AAGTGTGCAACATGAGTACCAACCACAAAGTTCAATAGANTANTCACCGATAGTCACGA
 CACGAACTNCCCTGNCGTATTCTCACCCNNAGAGGGCNATANCTCCCATTCTTAGCAG

TGTCAATATCCGTTCAACTGTCTAACTTCAAGANCTTCCCAGATTTTCGTTGACTT
GCTGTTCAATCGCACGCAATTCTCAGCAGTTACAGCTTGAAGTGGTAAAGTCAAAGC
GAAGGAATTCAACTTCGTTAACAGAGATCCTGCCTGTGTTGCGTGGTTCCAAGGATATTGT
GAAGGGCAGCGTGAAGCAAATGAGTCGAGTGTGGTTTCATGACACGGTGACGGCGAT
TGCTATCAATTGCCAACGGTATATTCTGTTCAAGGCAGCGGTGCAAGGACTTCACAG
TATGAAGGGCTTGACCATTGGGCTTCTGAACATTGGTCACAGTAGCCACAAACCTAC
CTGACTCATCCAAGAGATTGTCCGTAGTCAGTACCTGTCCACCCATTTCAGCATAAAATG
ACGTTCCGCAAAGATAAGAGAGGCAGTCCTCTGAAACAGCTTCTACTTCTGCATTGT
CCGCCACGATAGCTACCAATTAGAACAGACAATTGGCTAGCATTGAGTTGAAGGCACATT
CTACAGTGATGTTGAAGAGTTCACTTGACATACCCATTGAGCCACCCATTGACAGCTG
ACGCACGCGCGCTTCTGTTCTTCATGGCTGCTCAAAACCTCACGGTCTACAG
TCATACCAGCTTCTTCAGCAGTTCATCAATGAATTCAACTGGAACCCATAAGTATCA
TAGAGTTGAAGACATCTGAACCAGCAGCATAACAGATTGACCTTTCTTCAGTCTGCT
ACAATGCCTGGCAAAGTGTGACCTGAGTGAAGGGTACGGCAAATGATTCTCTCG
CTCTTAACGATTTCTCAATAAAAGTCACGTTCTCAAGCAGTCTGGTAGTAGCTTCC
ATGATTTTCAACAGTTGAAACGAGTTGTAAGGAAAGGCTCGTGATACCCATT
TGACCATGCATAGAACGACGACGGAGAACGACGAGAACATAACCAACGACCCCTCATTT
CCTGGAAGGGCACCACCGATGGCAAATGAAAGTGAACGGATGTGGTCAGCGATGACC
TTGAAGCTCATGGTCGCCATCTGGTCATAAAACCTTACAGACAATTCTCGACTTCA
CGGATAATCGGCATGAAGAGGTCCGTTCAAAGTGGCTTAGCCCTGGATAACGGCC
ACCAAACGCTCCAAACCCAGCGCCCGTATCAATGTTCTATGTGGCAATTCTGTATTG
CTACGAGGAACAGCAGGGCTCGCTTAATTGTGACAAAACGATGTTCCAGATTCAATA
TAACGGTCGTTCAATATCTCTGCAAGCAGGCGAAGACCGATATTCTGGTCAAAG
GCTTCCCCACGGTCAAAGAAGATTCTGTATCTGGTCCAGAAGGTCCGCACCGATTCC
CAGAAGTTGTCCTCAATTGGAATCAAGTGAACGGATCCACTCCACTCAATCCAGCGG
TTGTAAGAATCTTATCGTCTGGATAGTAGGTATGTAAAGTTTCAGCAGGGAAATCA
AACCATTCAGGGCTTGTCAAAGGCTCATAGCCAAAGTGAAGCTTCAGCGAACGTA
ATCCCCGATAGAGAACGAGTTCCCAACATTCAAACATGGTATGGTGACGCGCAGTCTTCC
CTAACGTTTCGATGTCGTTGGTACGGATAGCCTTTGGCATTGGTAATACGTGGATT
TCAGGGATAATGGTCCCGTCAAAGTATTCTTAAGGGTTGCTACCCAGAGTTGATCCAC
AAAAGAGTTGGTCATTACAGGAACCAAACATTACTGATGGTCTACTGAGTGACCTTG
GTCGCCAGAAATCAAGCCACATTGGCGTACTTGTGCACTAGATAGTTGTTCATATTG
TCTCCTATTCACTTGTAAATGTGATTGGCTTCCAGTATTCCACATAGTCATCGCG
ACACAGAGGGAAATGACTAGGTCTGCATAAGCGTCTCAAGAACCGTTACGGTATAGGTA
GAGGTCAGATGGAAGAGTTCCCTCTTAATTCCGCAATCAACTGATCGCGATCATCCAGC
GAATTGAAATTCAAATCCCAGATATTGCCCTCGATACGAAGACCTAGATTATCAAACCTC
ATACTTATCTGCCAAAAGGTCAACTTCTACGAATGACAAAACCTCGAGCCATCCGAAG
CTGAATCTAAAACGAGGAAGCAAGGTCAAGATTCTTACTGATCTGACTGACTTGTTC
ACCAGCCGCATCATAGATGGTAAAGTTGGGAAATCTTAAAAAATGATCCCTCACCTG
ATAGGCAATTCTCCCTGTCATCCTGATAGCGAAGCGTTCGCCTCCAAGACGAAACTT
TTGTTGACAAGAAATGTTTCATCAACACCTCCAAAATCAAAGACAAGCTCATATCA
CGAAGGGCGAAAACCGCGGTACCCACCTCATTCAATGAACCTGTCATTCTTGTCTT
ATGCAATTGTATGATTGAGTAGCATGACTTCCTAGCTTAGATGGCTCGCAGCACCGCCAT

TTCTCTGGACTAAGACAAGTGATATTCCGCCAAACTGGTCAATTACGGGTCAAGTCC
 TGCCTTCTGAGGGCACCAGGACTAGTATGGTGGACTAGCAAAGTGAACGTGCCTCG
 ATATCCACCCACGCTTAAGAGCAAGATAACCTGCTACAGGTGAGTCAATCCCTCCTGAC
 AACATGAGCATCCCTTACCTGAAGTTCCAATGGCAAACCACAGCCCCACGAATGGTT
 TCCATAAGAAAGATAGGCTGCTCTTCCACGAATCTCCACCCCTGAAGATTGATGTCCAGG
 ACTTTCCATTTGAACTTGCACATTGGAATGGCTTCCGAATACAGCCCCCTCCA

ORF Predictions:

ORF #	Start	End	Direction	Length
9	2812	3150	R	113 aa

[SEQ ID NO:] 3864214-9 ORF translation from 2812-3150,
 direction R
 VLMKTFLVKQKFRLGGERFAIKDDRGEIAYQVEGSFFKIPKTFTIYDAAGEQVSQISKEI
 LTLLPRFEIQLRDGSSFVIRKKLTFWRDKYEFDNLGLRIEGNIWDLNFKFAG*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864226
 Assembly Length: 2901bp

[SEQ ID NO:] 3864226 Strep Assembly -- Assembly
 id#3864226
 ATCGAATTTATTGACAGATTAGAAAAATAATGTTACATTATCCGCAGGTATCTTC
 GATACCAAATCTACATGAAGGGACGGGTATGAAACTTCTCATTATTAATTGGCTTAC
 TTCTACTCCTAGTCTTCTCTATTAGCATTGGGACCACTGATTTCATGGGGAAAGC
 TATTGATTTCGACCAGCAGACCTGGCTCCTCTTCAAGAGTCCGCTCCAAAGAACTA
 TCAGTATTCTCCTGACTGCCTCTAGTATGAGTATGGCAGGCCTCTCATGCAGACTATTA
 CCCAAAATCAGTTGCTGCACCGAGTACAGTTGGAACGACTGAAGCCGCCAAACTGGGAA
 TGGTGCTGAGCCTTTGTCTTCCATGGCTAGTCTGACCCAAAAGATGCTTCGCTT
 TTGTTTCATCCATCGTATTCCACCCCTTCTCCTAGCCTTATGACCATTAACTGTAA

AGGAAAGGTGGATGTTGCCTCTGATTGGGATCATCTATAGCGGGATTATCGGCTCAGTCA
CAGAAGTTATCGCCTATCGTTCAATCTGGTCAGAGTATGACTGCCTGGACCCAGGGCT
CCTCTCCATGATTCAAGACCCATCAGTATGAGTGGCTCTTCTTAGGCCTCATCATCCTGA
TAACCGTTGGAAATTATCCAAACCTCACCATCATGAATCTAGGGAAAGAAACCAGCG
AGAGTTGGGATTCCTACTCCCTACTTGAAAAACTGGCCCTTTCTGGTGGCGCTAA
CGACAAGCGTCACCATGATTACCGTGGGGGCCTACCATTCTCGGAGTTATCGTCCCA
ATCTTGTGCAAGCGCTATGGAGATAATCTAAGTCAAACCAAACATGGTGCAGTGG
TTGGTGCCAATCTAGTTCTGGCTTGCATATCCTATCCGAGTTCTGATTAGGCCATATG
AGTTGTCTGTCAGTCTTGCTAGGAATCATCGTAGTCTCGTCTTATCCTACTTCT
GGAGAGGGGGACGAAAAGATGCAGACTAAAGCAAACATACCAAGCTCTGGATTCTC
ATTATTCTGCCATCGGAGCTTGTCTCTACTTTGCCATCACTCACTTGTCA
TTTGCTTGGAAAGTTGCGTCCAAAAGATCATCGTTATCTTGTAGGCCATCGCAGT
GGGATTTCGACCATTAGTTTCAAACCTGACGGAAAATCGCTTCTGACGCTAGTATT
TTAGGAATTGAATCCTCTACGTCTACTACAAACCTACTACTGGTTTGAAAGCAAG
TTTCTTCAACTTGGCAAATCCCTATCTAGAATTCTAGTCTTACTTGTCCAGTCC
CTCTTCTTCTGCCCTACAAGGTTACTTGAAGACACTGATGAAGCAAGACCTGGTCTC
ATCCTGCTGATCTGCTAGCGCTCAGAAGTCTTCGAAATATCAGCACCTCCTTCAA
GTCCTAATGGATCCAAACGAATACGATAAAACTGCAAATAGTCTTGTGCCCTTCAA
CATCTCAACACTTCCATCCTAGCCATCGGTTCTGATCATCCTCGCTTGACAATCTT
TTCTTCGAAAAGCAGTCGTTAGATGTCTTGACCTGCAAAGAGAAACGGCTCAGATA
TTGGGACTCGATGTTGAAAAGAACAGAAAAGAGCTCCTCTGGGAATCGTGTGTTGACC
TCAACGGCCACTGCCCTGGTAGGACCTATGCCCTTCTCGGTTATGCTGGCAACCTC
ACCTACCTGATTGTCAAAGACTATCAGCACAAGTTACTCTTATAGTGGCATTCTGGTT
GGATTATTAGCTAACCTTGGGCAAGCCTGATTGAACGAGTCTTGCACTGGAAATT
CGTATCAGTATGATCATTGAGAGTGTGGTGGCTTATTCTTATCTTACTATAGG
AGGTCTCGTCAGTGAAAACATTGACAATCCATTCAAACACAGGATATTTGC
AAGGCATTTCGCTAAAGTCAGTCCCTAAAAACTGACTGCCATTGGTCAAATGGTG
CTGGAAAATCGACTCCCTCCATCATGAGCAGACTAACCAAGAAAGATCAGGGAGTTC
TCAGTATCAAAGGACGTGAAATCGAGAGCTGGAATTGCAAGAACTGGCTCAAGAACTAA
CCATCCTAAAACAGAAAATCAATTACCAAGCCAATTGACTGTTGAAGAACTGGTCAGTT
TTGGACGTTTCCCTACAGCCGAGGTGACTTAGATCAGAAGACTGGAAAAAATCCGAG
AAACTCTGAACTATTGGAACTGACCAACTAAAAGACCGCTACATCAATAGCCTGTCAG
GGGGCAACTCCAGCGCTTTATCGTATGGTACTGCCAGGATACGGACTTATCT
TGCTGGACGAAACACTCAACAATCTGATATCAAGCAAAGCGTCAGCATGATGCAGATT
TTCGACGACTGGTGGAGGAACCTGGCAAGACCATTATCATCGTCCCTCACGATATCAACA
TGGCCAGTCAGTATGAGATGAAATTGTCGCCTCAAGGACGCCAGGTCTTAGCAAGG
GAAGAACCGATCAAATCATGCAAGGCTGACCTACTCAGTCAACTTATGAGATTCCCATCA
CGCTAGCTGATATCAATGACAAAAGATCTGTATCTAGCTAGTAACATAAAAGCTCAA
GTTAGAGAACCTCAGTCTTAGTCAATAAGATCAAGAGACTCCCTAAATCGTTATCAC
ATTTTAAAAAGGAGAAATTATGAAAACATCCCTAAACTTATTCACTGCCCTAGTGGC
CAGCTTCTTGCTCCTACTTGG

ORF Predictions:

ORF #	Start	End	Direction	Length
8	1992	2744	F	251 aa

[SEQ ID NO:] 3864226-8 ORF translation from 1992-2744, direction F

VKLENIDKSIQKQDILQGISLKVSPQKLTAFIGPNGAGKSTLLSIMSRLTKKDQGVLSIK
GREIESWNSQELAQLTILKQKINYQAKLTVEELVSFGRFPYSRGRLRSEDWEKIRETLN
YLELTNLKDRYINSLSGGQLQRVFIAVMVLAQDTDIFILLDEPLNNLDIKQSVSMMQILRRL
VEELGKTIIIVLHDINMASQYADEIVAFKDQVFSKGRTDQIMQADLLSQLYEIPITLAD
INDKKICIYS*

Blastp and/or MPSearch Result:

Description:

ECFHUACD NCBI gi: 4143 - Escherichia coli. (fhuC, ferric enterobactin transporter ATPase, ABC type)

Assembly ID: 3864242

Assembly Length: 1930bp

[SEQ ID NO:] 3864242 Strep Assembly -- Assembly
id#3864242
CGANGGCCTTGATCTGGTGATGAAAAACAAGAATTGACTGCTGAAACTATCGTCATCAAC
ACTGGTGCTGTTCAAACGTCTGCCAACCTGGACTTGTACAAGCAAAACGTCTTT
GACTCAACAGGTATCCAAAGCTGGATAAATTGCCCTGAAAAACTTGGAGTCCTGGTGGC
GGAAATATCGGTCTTGAATTGCTGGCCTTACAATAAACTAGGAAGCAAGGTTACAGTC
CTAGATGCCTGGATACATTCTACCTCGTGCAGAACCTCCATCGCAGCTTGTCAA
CAATACCTGGAAGAAGACGGTATTGAATTGCTTCAAATATCCATACTACTGAAATTAA
AACGACGGTGACCAAGTGCTTGTGTAAGACGAAACTTACCGTTGACGCCCTT
CTCTACGCAACTGGACGCAAACCAATGTAGAACCTCAACTTGAACAAACGTTCTGGT
GACTAACTGAACGTGGCGCTATTAAAGTAGATAAACACTGTCAAACAAACGTTCTGGT
GTCTTGCAGTTGGAGATGTCAACGGTGGTCTTCAATTACTACATTCACTTGATGAC
TTCCGTGTTGTTACAGCTACCTGCTGGAGATGGCAGCTACACACTTGAGGACCGTCTC
AATGTACCAAATACTATGTTCATCACACCTGCACHTTCACAAGTTGGTTGACTGAAAGC

CAAGCAGCTGATTGAAACTTCCATACGCAGTGAAAGAAATCCCTGTTGCAGCCATGCCT
 CGTGGTCACGAAATGGAGACCTCGCGAGCTTCAAAGCTGTTAATACTGAAACA
 AAAGAAATTCTTGGTGCAGCATTCTCAGAAGGTTCTCAAGAAATCATCACATCATT
 ACTGTTGCTATGGACAACAAGATTCTTACACTTACACCAAAACAAATCTTCACTCAC
 CCAACCTTGGCTGAGAACTTGAATGACTTGGCGATTTAAGTTGAAATCTCATCTTAA
 CTGACAGCCCTCTTGGGCTGTTTACTTCTACGAAACACCAAATCTGCTTTCCCTC
 TTTTGTGATATAATAGAAACATGAACCTAAAAACTACTTGGGCCTCTTGCTGGCGTT
 TCTTCCCACTTCGTTTAAGCCGCTTGGACGTGAAAGTACGCTCCCAGGGAAAGTCGCC
 CTTCAATTGATAAAGATATTTACAAACCTAGCTAAGAACTACGAGATTGTCGTTGTC
 ACTGGAACAAATGGAAAACCTGACAACCTGCCTCACTGTCGGCATTTAAAAGAGGTT
 TATGGTCAGTTCTAACCAACCAAGCGGTGCCAACATGATTACAGGGATTGCAACAAACC
 TTCCCTAACAGCCAAATCTCTAAAAACTGGGAAAATATTGCCGTCTCGAAAATTGACG
 AAGCCAGTCTATCTCGTATCTGTTGACTATATCCAGCCTAGTCTTTGTCATTACTAAT
 ATCTCCGTGACCAGATGGACCCTTCGGTGAATCTATACTACCTATAACATGATATTG
 GATGCCATTGGAAAGTTCCAACGTGCTACTGTTCTCCTAACGGAGACAGTCCACTTT
 TACAAGCCAACATTCCAAACCTATAGAGTATTTGGTTTGACTTGGAAAAAGGACCA
 GCCCAACTGGCTCACTACAATACCGAAGGGATTCTGTCCCTGACTGCCAAGGCATCCTC
 AAATATGAGCATAATACCTATGCAAACCTGGGTGCCTATATCTGTGAGGGTTGATGT
 AACACGCTGATCTGACTATCGTTGACAAACTGGTTGAGTTGACCAACAATCGCTCT
 CGCTTGTCAAGACGCCAAGAATACGGTATCCAAATCGCGGGCTCTATAATATCTAT
 AACGCCCTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	376	1002	F	209 aa

[SEQ ID NO:] 3864242-6 ORF translation from 376-1002,
 direction F
 VLVVTEDETYRFDALLYATGRKPNVEPLQLENTDIELTERGAIKVVDKHCQTNVPGVFAVG
 DVNGGLQFTYISLDDFRVVYSYLAGDGSYTLEDRLNVPNTMFITPALSQVGLTESQAADL
 KLPYAVKEIPVAAMPRGHVNGDLRGAFKAVVNTETKEILGASIFSEGSQEIIINIIITVAMD
 NKIPYTYFTKQIFTHPTLAENLNDLFAI*

Blastp and/or MPSearch Result:

Description:

UNKNOWN DEHYDROGENASE A (EC 1.---). - ESCHERICHIA COLI.

Assembly ID: 3864254

Assembly Length: 2674bp

[SEQ ID NO:] 3864254 Strep Assembly -- Assembly
id#3864254
CTACTGTTGTTGATAAAGTCCTGAATCGGCTCTCCTGGTGGAGAGCTTTACTATTT
TCGAACCGACGATAACACCATCTGACACCGCATTGAAGCGTCCAGATTGGCTTACTAG
ATACACCAAAACCTGTCAAGACTGGGATGTCGGCCACTTGATGAAGTTGCGCCAAGTGCT
TGTCCTAACTCTGCATCGGTATTGCCTGATTTCCCTGTTACCCATTGATGGCAACGGCA
TAGACGAATCCCTCCGCCCTCAATCAACTCTTCTGGCGCTCAATTCTGTGGTCAAG
CTTACTAAAGGAATCAAGGCATATCTGTATCTGCCAAAATGGTCTACAAAGTTGGCA
TGTTCATGAGGCAGGTCTGGATAATCAAGCCCTCACAGCTGTATCAGCCAGATTTG
ACAAAGTTCTCCACACCGTACTGAAAGAGGGGTTGAAGTAGGTATGACCGAGTGG
ATCTCTGTTCAATGGTTCAAGGTTCAACTAAAGCCTGGTAGAGGTCCGTGGCT
AAACTGCGCAAGCCAGCTTCTCAATAACAGGTCCATCTGCAACAGGGTCTGAAAAGGGA
ATACCCACTCAATAGCAGAGACACCCAAATCTCTAAAAAGTGAATTGTTCAAG
CCGTCCAAACCTTTCTGGTCAACAGGCATGATATAAGGAACGAAAATTCTTTCCA
GTTGCTTTATAGCATTCAATTCTGTTAGTGTCTAGGCATGAGCTCTCCCTTCT
TGCTGCATCTGCTTCCAAGCGGTCTTGACTTGAACCACATCCTGCCCCACGACCTGA
TAGGCAGACAATCATAGACTTTCTGGTCAAGTTCTGGCAATTCAACCGCAAAGC
GATAGCATGGCTAGATTCCAAGCTGGATAATCCCTCACACGAGACAGAGTTGGAA
TCCTTCCAAGGCTTCTCGTCTGTCACAGGGACATAGCTGGCACGTTAATATCGTGGTA
GTGAGAATGCTCTGGACCGATACCAGGATAGTCCAAACCTGCTGAGATAGAGAAGGCTC
AAGAATTGACCATGGCATCTGGAGCACATCCATGAGGGAACCGTGAAGGACACCTGG
ACGACCCCTGGTCAAGGTAGCTCGTGGTCTCGTATCCACACCAAGTCCAGCCGCTC
AGCTCCATACATGGCTACAGACTCATCTACAAAGGGATGGAAGAGCCAAATAGCATT
AGATCCACCAACACAGGCTACTAGGGCATCGGCAGATTGACCTGTCAATCGCG
ATACTGTTGTTAGCTCGGACCGATGACACTTGGAAAGTCACGAACGATTCTGGAAA
TGGATGAGGCCCAAGGCAGAACCAAGGGATATAGTGGTATCGTCGATATTAGCCACCA
TGAACGAAGGGCTGCATTGACCGCATCCTGAACACCGCGAACCATCTGCACTGCC
AACCTTAGCTCCAAAAGCTCCATACGGAACACATTGAGGGCTGGCGTTGACATCTC
CTCACCCATGTAGATGGTACATTCCATGTTAAAGAGGGCCGAGCAGTTGCAGTTGCC
ACCGTGCTGACCAGCACCGTTCTGCATAATTCTTACCCATGCGTTGGCAAG
CCAAACTGTCCTAAGGCATTGTTAATCTGTTGGCTCTGTATGGTTAAGGTCTCC
TTTGAGATAATCTGGCTCCGCCGATATGCTGGTCAAGTTTGCAGTAAAGAGG
AGTTTCACGTCCTACGTACTGGCGCAAGAGTTGGTTAACCTCTGGAAACTTGGTC
TGCGTACTTCACGGTAGGCCTCTCCAACCTCCAAAAGTGCCTGTCATCAATGTTCTGG
GACAAAACGTCCGCCGAATTTCCTGAAATCCATCTTATTGCTGCTGATATGCCAT
GCTTACCCCTCTATAAAATCTCTAATCTTCTGATCTTTGCCATCTGCTCCA

CTCCGCTCGATACATCTACTGCATAGGGAGTAAATGTTGAATTGCTTTACTACATTAT
 CTTCATTAAGGCCACCTGCGATAAAGAAGGGCTGTGCTAGTCCAGTCGTATCCAGTTGAC
 CCCAATCAAAGGACTGGCCACTTCCTGCCACAGGGGCATCAAAGAGTAGATAATCTGCCT
 GAGAATTGGGGACATGCCATTCCATCTACCTGCACAGCCTGAATACTGGCACAGGCA
 AATTCTCAAATAATCATCTGCCACCTGACCGTGAACCTGAACCAAGTCCAAGCCAACCTT
 TGTCAATCGCTTCCAGCAGTTCTACCCGACTTGGTAAACAAATACTCCAACCTTTCA
 CATCTGCAGGAATAAGCTTGCCAACTCAGCTGCCTCTTCTAAAGTCACCTGTCTTAC
 TAGGTGCAAAGACAAAACCGATATAGTCGGCTCTGCTGAAACGGCTGTTCCACCGCTT
 CTTGGTCGATAGTCACAAATTAAACCTTGTCAATCTGCAACTCCTGATTCTCTGG
 GCCACATTTCTGCCTGCATAAGAGCTGCCCTACCAAAATTCCGTTAAAGTATGGGCT
 AGTCGTTCCGCATCCTGCCCTGTGAAAATGGCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	117	833	R	239 aa

[SEQ ID NO:] 3864254-6 ORF translation from 117-833,
 direction R
 VGTRMWFKSKTAWKQMQRREKLMPKTLTEKLNAIKATGKGIFVPYIMAGDHEKGLDGLA
 ETIHFLLEDLGVSIAEVGIPFSDPVADGPVIEEAGLRSLAHGTSTQALVETLKTIEIPL
 VIMTYFNPLFQYGVENFVKDLADTAVKGLIIPDLPHEHANFVEPFLADTDIALIPLVSLT
 TGIERQKELIEGAEGFVYAVAINGVTGKSGNYRCRFGQALGATSSSGRHPSLDRFWCI*

Blastp and/or MPSearch Result:

Description:

TRYPTOPHAN SYNTHASE ALPHA CHAIN (EC 4.2.1.20). - LACTOCOCCUS
 LACTIS (SUBSP. LAC TIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3864296

Assembly Length: 3074bp

[SEQ ID NO:] 3864296 Strep Assembly -- Assembly
 id#3864296

CCAACATTACATGTTCCAATTTCTGGTTGGCTGTTGAGTTAACAAATACATAA
TCTACACCTGTCAAAACGATGAAGAGGCTGCATCAACCAATTCTGCCAACGTTGGAA
GCGAAGTCTTATCAATAACCGCTTCGACACCAGTCAAATGTCATTGTTCTTGACG
ACGGGAATACGCCACCACCTGCAGCTACGACGACTTGACCATTATTAAGAGTACGG
ATGGTTCAATTCTTGATATCAACAGGTTGGTGAGGCAACGACCTTACGCCAGCCA
CGGCCAGCATCTCCTGAAAGTCGCTCGCTTTGGCTCTGCTTTGCTTCTCT
TCTGAATAGAAAGGACCGATTGGTTACTCAAGTTAACAAAAGCCGGATCATTATCT
ACGACAACCTGCGTTACAACAGAACATTTTCGATGCCTCATCCAAGAGAGCA
TTTGCAAAGCATTTCACCAACAGAAACCGATGCTACCTCTGTACAGCGACAAGTGAG
TCGAATGGGAAGGCAGGGTCTTCAGAGTTCTGATGCCAAATGTTGGAGCAAGAGA
TTCCAACCTGAGGTCCCATTACCGTGAGTTGATAATCAAATCATCTCCATTAAATCC
AATTTACAAGATGCTTAGCTGTTCAACTAAAGCTCCTTGTTGAGCCCTTGCTGAT
GGGTAGAAGAAAGAACGATCGATTCCCTCCAAAGCTACTACAATTACGATTGCCATA
AATTCTCCTTATCACACTCAATAGAATGCGTTAGATTCAATTAAATGATTTTCACA
TATTTATAAGAAATAATAGATTACCATTATATAAAAGAGGACGGACTAAAGCTATTAG
TCGCAGCCCTCATAGCTGTTGAGACGGTTATTATCTAAAATTATACTTTAGGAATAT
AAAGGTTACCAAGTGTAGCAGCCATAACAGCTTGATAGTGTGCATACGGTTCTGCTT
GATCGAAGTGGCGAGCGTACTTGCTGCGGAAGACTTCGTCGTTACTCCATTCTCTA
CACCAAATTTCAGCAACGTCTTACCATAAACAGTGTGAGTATCGTGAATGCTGGCA
AGCAGTGTAGGAAGATCAAGTTTCATTGCGCTGTTTAACTAAGTCATATTGACTT
GGTAAGGTTAAGAACGAGCTACACGTTCTGCAATTGCTCTTCACCCATTGATACCC
AACGTCTGTAAGAACGTCTGCATCTTAACGCTTCATCAGCATCTCAGTGTGATGA
AACATGTGCGCCACTTCTTAGCAAATCCTCTGCAATTCAACGATTCTTTCTG
GGAAGAGTTCTTGAGAAGATGTGAAACATTGACACCAAGGATAGCACCTGTTACGA
GCAAGCTGTTGGCAACGTTGTTACGTCCATCACCACAGTACCAATGTCAGCCTTCCA
AGCGACCGAAGTTCTGAAACAGTCAAGTAGTCAGCGAGCATTGAGTTGGGTGCCATT
CGTCAGTTAGACCGTCCATACTGGAACGCCTGAGAATTGCAATTCTCAACCATAA
CGTGGCTGAATCCGCGAATTCAATCCGTCAAACATACGTCCAATACCTTAGCAGTA
TCTTCAGTAGATTCTTTTACCCAACGTGAATATCATTGCTCCGAGGTATTCTGGGTGA
GCACCAAGGTCGATGCCGAGTTGAAAGGCTGCACGAGTACGAGTAGATGTTTTCA
AATAGGAGAGCGATATTCTGCCAGCAAGGTAGTGGTGTGAATATTGCGTTTTCAAA
TCTTCAGTGAGCTGAAAGACCAATAAGGTATTCTAACTCTGACGGGTAAAGTCTTT
TCTGCTAAGAACGCTGCGTCTTGAATACTGAATTGTCATTATTATTCCTCTTCT
ATTTTTACATTCTATTGACGAATGCCAACAGCGATTACACTTCTCACGTTCAAAT
GGCATAGACATAACAGAGGTCCACCACGGCCCCAACCAATTCACTCCGCGAATCTTA
ATCAAGCGAAGCCGTATTCTCCAAATCTATTGGTCACGGTATTGCGGTACAAACA
ACTACCACACCAGGTGCGATGGTCAAAGTGTAGAACCGTCGCTCCATTGTTACACCAAGGTT
GCTGCTACGATATTGCCACCACCGCAACGAATCAAATGAACTTTCTACACCAAGGTT
TGAGCAAGAAGTTCAGCTAACGTACACCTTCTTCAACGATTAAAGTTCTGGTTTCG
TAAGTAACTGAGTAAACGTGAAGGTGGCCTTCGATTCTGGGTGAATAGTGAACCTGTC
ATAGTCTACCATAGTGAAGACAGTATCCAAGTGCATGAATTACGGTTAGCAAATTC
AAAGGCCAAAACCTTCTGAAAGCCAACATTTCCTGAAAGATGTTGACCAAAAGTTTC
GATAGAAGCTGCGTCTGTACGTTGAGAGATACCTACTGCAAGGACGTCTTAGAAAGAAC

TAGCCTCGTCTCCACCTTCGATACCGTATCTTCTTCACGGTTGAGACCAAATCCACTT
 TTCCGCCATAGATTGGGTGGTATTTGAAGATATACTTACCGTAGAGTGTTCACGGTTAC
 GAGTGTCTGCAAACATGTGGTTAACGATACGGCGTTCCAATTGTTGCAAATGGGTGGC
 GAGTGAAATAGAGGTTGGCATCGGGTCAATTGCAAATGGATAATCTGATTCAACTAAGT
 CAGTTAGATCTTAGCTTCGTCAGGAATTCTGGCAATTCAACTTTGAATCCCAGCCA
 TTGTTTTTCAACCAATTCTTGTTGTCCTGATGCCGTGAAGCAATTCAACGAATAGCAA
 CCTTGGTTTGACGATCACGGATGTTGGCTCGTCAAGTATTCTCGATAAAATTGATCGC
 GGATTCTGGAGAAGTCCAATGAATCCAGCAGCGAGTTGTTCTACCTCCAGAACCGATTA
 TCTGCTGTTGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	944	1777	R	278 aa
10	2323	2694	R	124 aa

[SEQ ID NO:] 3864296-7 ORF translation from 944-1777, direction R
 VQPLQLRLSTLVLTQNTSEQMIFSWVKKNLLKILLKYWDVCLTGLNSADSANVMVEELAE
 FSGVPVWNGLTDEWHPTQMLADYLTVQENFGRLEGLTLVYCGDGRNNVANSLLVTGAILG
 VNVHIFSPKELFPEKEIVELAEGFAKESGAHVLITEDADEAVKDADVLYTDVWVSMGEED
 KFAERVALLKPYQVNMDLVKKAGNENLIFHCLPAFHDTHTVYGDVAEKFGVEEMEVTD
 EVFRSKYARHFDQAENRMHTIKAVMAATLGNLYIPKV*

Blastp and/or MPSearch Result:

Description:

ornithine carbamoyltransferase (arcB) homolog - Haemophilus influenzae (strain Rd KW20)

[SEQ ID NO:] 3864296-10 ORF translation from 2323-2694, direction R
 VKHSTVSISSNTTQSMAEKWIWSTTVKKIRVSKVETRLVLSKDVLAVGISQRTDAASIEK
 LLVNIFKKNVGFKKVLAFEFANRKFMHLDFTVFTMVDYDKFTIHPEIEGDLHVLLSYLRK
 RKT*

Blastp and/or MPSearch Result:

Description:

STREPTOCOCCAL ACID GLYCOPROTEIN. - STREPTOCOCCUS PYOGENES.

Assembly ID: 3864300

Assembly Length: 3205bp

[SEQ ID NO:] 3864300 Strep Assembly -- Assembly
id#3864300
GGGGCAAAGCAAAGACTCAAATAGCTAGAACCTACTAAAAAGATGCTGAAATTCT
TATATTGATGAAGCCACTGCTAATCTTGATGCGGATTCTGAGTATGCGATTATCAGTAG
CCTCTATTCTGTATTAAGGAGAACGGTTGTGATTATAGCGCATAGTTGTCACGGT
AAAAGATGTGGATTGTATTTCTCTAGAGGAGGGAAAATCACTGGCTCAGGAACACTCA
TAAGGAACACTGGAAAATCATGAGCGTTATGCTCGTTGTGCAGGAGCAAATGATAGA
GTGAAGTGTCTTGAGATTACCACTTTATAGTCTATTAAAGGGAGCAGGAAAAACTCC
CTTTTATATAGTTGAAACTATAACTAGCTCTGAAAAGAAGAAAATGAGTTGATGAAA
ATAAGTGGTACAATAGTTACTATAGATTGGAGGTATTGTATGAGCAAGGAATTACACAT
TAACACAATTGGCCCAGGCGGGTATTAAGTCAGATGAAGCGACAGGTGCATTGGTGAC
ACCGCTTCATTTCAACGACCTATCAGCATCCAGAGTTGGTCGATCTACTGGTTGA
CTATACGCGCACTAAAATCCAACTCGTAGTAAGGCTGAGGAAGTCTGGCGCTATTGA
GTCAGCAGACTATGCCCTAGCGACTAGCTCAGGGATGTCAGCTATTGACTGCCCTTAG
CGTCTTCCAGTAGGAAGTAAGGTCTGGCAGTGCAGTGATCTTACGGTGGTCTTTCG
CTGGGTTAAACCAAGTGGAGCAGGGAGGTGCTTCCATTAACTATGCCAATAACA
GAAAGGAAGAGTTGATTGCCGGAGTTAGGAAAAGGATGTGGATGTTCTCTATATCGGAAA
ACCCCAACCAATCCCTGATGTTGAAATTGATATCGAAAAACTAGCAAAATTGGCTCAT
GCTAAGGGTGCCAAAGTGGTGGACAATACCTCTATAGCCCTATCTACCAACGTCCG
ATTGAAGATAGAGCAGATATCGTTCTCCATTCAAGCAACCAAGTATCTAGCAGGCCACAAT
GATGTCTGGCTGGAGTGGTTGTGACCAATAGTTAGAACTATACGAGAAGCTTTTAC
AATCTCAATACAACAGGGGCAGTCTGTCTCCATTGACAGCTACCAGTGTCTCGTGGT
CTCAAGACCTTGTCTTCGTATGGAGCGTTCAACAGCTAACGCCAAGAAGTGGTGGC
TTTTGAAGGATTCTCCAGCAGTTAAGGAAGTCTCTACACTGGTCGTGGAGGCATGATT
TCCCTTAAAGTAGCCGATGAAACACGCATTCTCATATTGAAACAGTCTCAAGGTCTC
TCTTTGCGAAAGTTGGGGAGTGGAAAGTCTTATTACTTACCAACGACTCAAAC
CATGCTGATATTCCAGCAGAAGTACGCCATTCTATGGTTGACAGATGACCTTGC
TTGTCTATTGGGATTGAGGATGCTAGAGATTGATTGAGATTGCGAGATTGCGCAAGC
GGATAAGACAAAGATGGAAAATATGATTTACAAGCCTGCCAACCGTTAGGGCACCA
TACCTATAAAATGGAAAGAAACAGAAACGGATAGTGAAGTTCTACCAAGCTGGATAGCG
TATGGACTTGTGGCTTGCCTGAAATCCGCCAACCGTGCAAACCTACGCAGACCAACT
GGTTTATGGTTATACCTATGCCAGTGAAGACTTAAGGAAGTTCAAAAGTGGGAAGC

TACACAATACGGTTACAACCTTGACAAAGAGGCTCTGTCTTATCGAGGGTGTGGTACC
 AGCCATCTCAACAGCTATTCAAACCTTACAAAAGAAGGCAGGCAGGTTAATTAACAC
 GCCTGTCTACCCACCCCTTGCTCGCAGTGTCAAGTTGAATAATCGTAGATTGATTACTAA
 TTCCTTAGTGGAAAAGGATGGCTGTTGAGATTGACTTGACCAACTGAAAAGGATT
 GGTGGAAGAGGAGGTTAAACTCTATATTCTTGCAACCCTCACAATCCTGGTGGACGTGT
 TTGGGAAAAGAAGTGTGGAGAAGATTGCCAACTTGCCAAAACACGGTGTGTT
 AGTTTCGGATGAGATTCACCAAGATTGACCCTCTTGGTCACAAACACCCAGTCTTCAA
 TACCATCAATCCTGCCTCAAAAATTTGCTATCGTCTTGAGCAGTGCCACTAAAACATT
 TAATATTGCTGGAACAAAAATTCCATGCAGTCATTGAAAATCCTAAGTTGAGACTAGC
 TTTCCAGAAACGCCTGTTGGCCAATAATCAGCATGAAATTTCAGGCTTGGTTATTGGC
 GACAGAAGCTGCCTATAGATACGGTAAAGATTGGCTAGAGGAACCTAAGCAAGTCTTGA
 AGACCACATCAATTGATGTGGTGGATCTATTGGAAAAGAGACTAAAATCAAGGTATG
 AAACCGCAAGGTACCTACTTGATTGGCTGACTTTCAAGCCTATGACCTGACTGATGAA
 ACATTGCAAGAGTTGAGAAATGAAGCCAAGGTTATCCTCAACCCTGGTTGGATT
 GGAGAGGAAGGAAGTCTCCATTCCCGCATCAAGATTGTTAGCTATGCCAAATCTCTGTT
 GCAAGAAGTCTGTCAAGCGGATTGTGGCTACTTTGCCAACGTTAAAATCCAGCCTCT
 AGGAGAAAAGTCTCCTAGAAGGCTATTTCATAGGCGAAAATATGGTATAATAAACAGA
 TAAGGTAAAGGTGAAAATATGGCTAAATTGATTCCGGGGAAAGTCGTATCGAAGGTGTT
 GCCCTTATGAAACTGGTAAGGTTGATATCATCAAGGAAAAGAACATCGGCTCTACGCT
 CGCGTTGCAAAAGAAGAACTGCGCTATAGTTAGAGGATGATTGGTTTTGTGCCTGT
 GATTCTTTCAAAAGAGGGCTACTGTGTGCATTGGCAGCGCTAGAGCATTCTGAAA
 AATGATGAGCGTGGTCAGGAAATCTGTGGAGTCAGGAAAGGTATGAAGAAAAAGAG
 GCCGTTGAAACCAAGGTGACCTGGTGGCAAGTTTGAAATCGAATTCTCCGAAA
 TCAGAATGCGCCTATGAGTTATCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
9	2479	2823	F	115 aa

[SEQ ID NO:] 3864300-9 ORF translation from 2479-2823,
 direction F

VVDLFGKETKIKVMKPQGTYLIWLDFSSAYDLTDETLQELLRNEAKVILNRGLDFGEEGSL
 HSRIKIVSYAQISVARSLSADCYFCQTLKIQPSRRKVFLEGYFHRRKYGIINR*

Blastp and/or MPSearch Result:

Description:

PUTATIVE AMINOTRANSFERASE B (EC 2.6.1.-) (FRAGMENT). -
BACILLUS SUBTILIS.

Assembly ID: 3864312
Assembly Length: 1665bp

[SEQ ID NO: 3864312 Strep Assembly -- Assembly
id#3864312
AATTGATGGCGCATATAGGCTTCCATGGACCTTGTAGAGTCTTGCTGCTTCT
AGCTCCTCAAGTAAATCTGCTAAACTCATCTAAAACCTCTGCCCAACAAATGGTGC
TGAAAGGCATACACAGTCGCCTGGGTACGATCGCTGACTTCAAGTTGGCAAGAATATTG
GACACGTGGTCTTGACCGTCTTGAGAGAGATAAAAGAGGTACCTGCGATGCGCTGATT
TCGTAGCCCTGGCGATGAGTTGGAGAACATCTCGCTCACGCGAGTCATTCTCATGA
AGTTCCATATGATTGCGGTGGTATTCAACCTTCTTGCTAACCTCTTGCTCAATGCCAGC
TCGCCAGCAGCTACCTTACTGACGGCATGAAGCAATTCTGCAGTACAAGTCTTGAGC
ATATAGCCTTGGCACCAGCATCTAACAGACTGGCATGATTTTCATTGTCCAATAAGAG
GTCACAATCAAATCTTGGCTTCAGGCCATTCTTAAGGATTGCTAACGGTGCAGTCATC
CCATTCATCTCAGGCATGACAATATCCATGACAATGACATCTGGACGCAGTCCAAGGCC
AAGTCAATCCCTTGAGACCCGTTGGACGCCTCACCCACAACCTCTACATCGTCTTGAGG
TCAAAGTAGCTTTCAAGCCAATCGGACCATTCTGGTCATCTACTAGTAAAATTTC
ATCTTACTCCTTATCATTCTTATCTAACAGGGAAATACGGATATCAACTGCCAGCCC
TTGCTTGGAGCTGTTAATAACTGAACCGTCCCTGCCATATCTCAACCGCTCCTTGAT
ATTTCGCAGTCCATAACTCAAGTCGCTAACGCTCCTAACCGAAACCAATCCATTGTC
CACCACTTCAGTTGCAATTCAACATCTGCTGATAGAGGTAGACATCTAGGCAAGATGC
CTGGGCATGGCGGAGCGTATTGCTAACACTCTGCAGGATAACGGAAAGATATGCTCCTC
GATTTCTTATCGGCAATTCTGCTATTCTGCTTGAGACTAACCTAACGATCACTCTTG
TCCTCAAGCTTTAAGAGAATCTGAATCCCTCTATCAAGCTCTGCTCCAGTTCA
ACTGGTCGCAAATGCAAGAGCAAACCGCAAATCCCTCTGGCAGTTCTAAAGACT
GTGACACTCTGCAACTGGATCTGCATCTTCTATCCAATTCAAAGCCTGCTGACTG
ATACCCGATAAAATCATGTGGGCCGAAACAACTCCTGACTGACTGTATCGTCAAATCC
CGAGCAATTGCTTCCGTTCTCGATGATTTCTCTTGAGCAAGGCTATGATT
TCAGCTTTGAAGAGCTCTGCTAAAAGGTTAAGTTACCTGATAAGGACTTGAAACTG
GCATCCAAATCTGGATCTGCAACCTGAACCAACTCTGCCCTGCCAATAACGCTTGAGA
TTAGCCTGCATTCTTAGAGAAAGCTCTCGATCCCTGCCAAACAGGGCTAACAGAGA
CAGGTTATGGACATGCTGAAAACCAACAATAAAAGACAAATTCTGTTTCGACA
TCGTGCAAAAGATAGACCGAGTCAGTAAATCAAGTATTCCAGCAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	736	906	R	57 aa

[SEQ ID NO: 3864312-7] 3864312-7 ORF translation from 736-906, direction R

VVDNGIGFRLGSLLDSYGLRNIKERVEDMAGTVQLLTAPKQGLAVDIRIPLLDKE*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3864336
Assembly Length: 2532bp

[SEQ ID NO: 3864336] 3864336 Strep Assembly -- Assembly
id#3864336

CTGAGTGAAGAAAAGTACCACCACGAGAAATGATGTCTCCTACTGAAGCTGCATCTAGGG
GATGAATTCAACCGGCAACCATACCAGCATATCCGTATAGATAACAAACACTTCCATT
CTTCTGAAATTGCTTGACGAACAACGTGCACGGATAGCAGCGTTCATACCAGGTGCGTCTC
CGCCACTAGTCAAAACAGCAATACGTTCATATTGGTTATGCTCCTTTCTTTAAC
TTCTTCTTGATTATATCACATTGATTAAAATTCTCTATTCCGTATTTAGCG
ATAAACGTTTCATAACGATTCAATTCAATTCCCTCTAATTGATTGGATGACTGGAT
AAAATGATGGGGAGAAACGATGGTTCTGTTCTTACCGGATGATGACTGGAT
TGGCCTTAAATTGTTCTAAACGATGAAATTCTTGATCCGATTGATGATTTCAC
CTGTATCCAAAAGCGTTCAGCAACTGCTCTTATTCTGTGCAATCATTGCAAACG
GCCATCACGTGATTGTATTTCTTACATAGTAGAAGGCTCCCTTTATTCCTG
TCCAACCTGACGATATAAGTCTGAAAAGAGAGTGACATCCAATTCTTACTATC
TGCCTGTAAGAAGGCCATATTTCACCCCTTTGGTACGAATCACTTTATTCTGAAC
TTCAACAAAATAATAGCATAGCTATTCTGACAAATTCCGATTGGGTAATCGGGTA
AATAGCCTTACTTGCAATAGCTGTAGNGGATGTGCTGACACCTATCCCTAAAGCTC
TTGTTCCATATAAAATTCTGTTCCGTCCAATCTCCGATTGCCAACTATAAAT
AGCATCTCCAAACAAACTCCCAACTCTTACAAATTCAAATAGATTAGCTAAGTTATT
AAATACTTTTGACGATTTTCAAATGAATCGAAAAGACCAACTTTACCAAAGGTT
TAGCAGAGGAAGTTCAGATAATTCTCAGGTAATTAGCTATAAAATCTCAATGTTAGA

ATAAGGTCTATGTTCAATAATCAAAGCGCCAAGTCCTGCTGAGCCCTTAATCGATT
 CAAACCTATATAGATAGACTTGTGGCAATTATCGTGTAGGGAAATAGTATTGATGGA
 TAGAGAGGCTACTTCAAAACCTGCTCAAGTCATCTATTAAGTAATCACTGTTGGAATA
 ATTTAACATGACCTGATAAAAATGGCTGGATAATGCCTTGAAATAAGCCAACCTGGAA
 GGCAAGGCTGAGTAGGCGTAGGCATGAGATCTATTAAATCCATAACCTGCAAACCTCTC
 CATAACATCAAAACCTGCTCTGATTTCCGAGTATGCCCTGCTCTATGGAGCCTG
 AATAAGGAAGCCCTCATCTCATGCATAGCAGAGGCATCCTTTACCCATAGCTGACG
 CAAAATATCGGCCTCCCAAGACTAAATCCAGCAAATCGCTGAGCAACCTGCATAACCTG
 CTCCTGATAGAGCATAATGCCATAAGTTGGAGCCAAATATCCTCCAGAGCTGAATCTAG
 AACAGTCACCTCTCCTGCCATGCTCCTGCCACAAAATTATTGATGTAGTCACTTGC
 ACCTGGTCGATTTAGAGAAGTAGTTGCTACGACATCTCAGAAACAGACTGGTTAACACG
 TTTGAGCAAGCGAATGGCACCAGGTTGCTCAAATTGAAAGATAACCTTTGTATTCAGA
 GGCAAATAAATCTAACGTTCTTGTCTCAAATCTATTCTCAATTAAAGGTGAAT
 ACCTTCTGTTCAGCAAGCAACTCTGCATCTCTGGACAAAGGTCAAATTCTGAGTC
 CAGAAAGTCCATCTCAAAAGTCCGCTAGCCTCAACTCCATGAGCATACTGAGTCAG
 TGGAAATTTCATCACCAACTTTAGAGGAATGTAGTTGGTAAATCTGGTCACTAATTAC
 AACACCAGCCGCATGGACAGAGGTTGCCTGGATAGCCCTCTATCTGCAAGCAATCTC
 AAAAGCTTTGGTATTCTAACTTACTATTGATTGGCTGACGAAACTGGAGATTGCCCT
 CATAGGCCGACTTAAGATTGTACGAAAATGATTCTTAGTAATTGAGATAATTCT
 ACTCTGGCACACCAAAGCGTTCAAGACATCTCGAAGAGCTTGGCTCCAAAGGTTG
 AAAAGTAACGATTGTGCGCATGTTACTACCATAATTACCAACATATCTGATAA
 AATCTGGACGATAAATCTGGGATATCAATATCAGGCATGGTATAGCGTTCAC
 GATTAAGAAAGCGTTCAAAATCAGATTCTACTGGGTCAATCCCCGTGATGTCTA
 AGGCATAAGAAACCAAATGCCTACTGCAGAACCCCTCCATTCCATATAATAGCCAT
 TCGATCGTCCAA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	295	2232	R	646 aa

[SEQ ID NO:] 3864336-6 ORF translation from 295-2232,
 direction R
 VCQSMNYLQLLRKSVFVTILSRPMRAISSFVSQINSKLEYQKAFEIACKIEGYPRQTSVH
 AAGVVISDQDLNYIPLKYGDEIPLTQYDAHVEASGLLKMDFLGLRNLTQVQKMQUELLA
 ETEGIHLKIEEIDLEDKETLDFASGNKGIFQFEQPGAIRLLKRVQPVCFEDVVATTSL
 NRPGASDYINNFVARKHGQEEVTVLDSELDILAPTYGIMLYQEQMVAQRFAGFSLGK
 ADILRRAMGKKDASAMHEMRASFIQGSIEAGHTAEKSEQVFDVMEKFAGYGFNRSHAYAY
 SALAFQLAYFKTHYPAIFYQVMLNYSNSDYLIDALEAGFEVASLSINTIPIYHDKIANKSI
 YIGLKSIGLSDLALWIEHRPYSNIEDFIAKLPENYLKLPPLLEPLVKVGLFDSFEKNR

QKVFNNLANLFEFVKELGSLFGDAIYSWQESEDWTEQEKFYMEQELLGIGVSIHXLQAI
SKAIYPITPIGNLSENSYAIILVEVQKIKVIRTKKGENMAFLQADDSKKLDVTLFSDLY
RQVGQEIKEGAFYYVKGKIQS RDGRLQMI AQEI REAVAERFWI QVKNHESDQEISRILEQ
FKGPIPVIIRYEEEQKTIVSPHHFVAKSNEEKLNEIVMKTIYR*

Blastp and/or MPSearch Result:

Description:

DNA POLYMERASE III, ALPHA CHAIN (EC 2.7.7.7). - ESCHERICHIA COLI.

Assembly ID: 3864344

Assembly Length: 2244bp

[SEQ ID NO:] 3864344 Strep Assembly -- Assembly
id#3864344
GTTAACCTAGAGTAATCATTTCACACAGTTACGGATTCTTAGCACGAGCTTCAG
TTGTCACGATTGATTGTTGATCAAAAGGTAGTTGTCAAATCGCGAACATTGCTTAC
GTTGTGAGCTAGTGCGTCCTAGTTACGGTAAGCCATGTATTCCCTCTTATTTATCTT
TAATCCAAGACCCAATCAATGAGTTGAGTTCACTTCTCAAACACTTGCCTCCAAG
ATTCGTACTTTCATCATCTGCTTCAGATTTCTGTCAAATCATGCACAGTATTGAT
ACCGGCACGTTTAAACAGTTGATGAACGCACAGACAAGTCCAGTTCTCAATCGTACG
ATCTAAAATACGGTGTCAAGATTCAAGTCTGTTGACTTCAGCTTCACTTCAGTTGACTTAGC
AATCTCAGTAAGATTGAAACAAATCAAGATGTTCTGTCAAACATCGTGTGAAAGCCC
TAAAGCATCTTCTGGAATAATTGTTCAAGATTCAAGGGTTAATTGTCGAA
ACCATCATTGCTACCTACACGAGCAGGTTCCACTTGATAGTTGACTTTGTAAGGGTGT
ATAAAATAGAATCTACAGCAAGTGTCAACTGGTGCATTATCCTTTATTTCATCAGC
AGGTACATATCCACGACCCTGTTAACAGTCATAGTCGCTTTAGAGAAGAACCTTCACC
AATTGTAAGAGATAATGATCTGGATTACAATTCAATATCGCTATCTGTCAAATGTC
ACCAGCTGTTACTTCAGCAGGACCTTCAACATCCAGTTGATGATTTCGTCTCAAC
GTACGATTTCACTGCAATTCTTAAATGTTCAAGATGATTGATCAGTCTCACGAAC
ACCTGGAACGTGTCAAACACTCATGTAACACACCACATGTTGATAGATGTCACAGCTGC
TCCTGGTAGAGAAGCTAGAAGTACACGACGAAGAGAGTTACCAAGAGTTGACCGTAGCC
ACGTTCAAGTGGTTGATTACAAACTGCCATAATCTTATTTCATCAATTGGTTAT
ATTGGTTTCAAACTCGATCTTAGTTACTCCCTCTAAACGAAAAGCAGTGTAAATG
CGATGATTATACACGGCGACGTTGGAGGACGAGCACCATTGTGTGGACTGGAGTCAC
ATCACGAATTGCTGTACTTCAAGACCAGCGGCAGCAAGCCACGAATAGCTGACTCACG
ACCAGAACCTGGACCTTACAGTAACTTCAACTGATTAAAGACCGTGTCTGTGCAGA

TTTAGCAGCAGCTTCAGAAGCCATTGAGCAGCGAATGGTGTACATTACGAGAACCTT
 GAAACCAAGAGCACCAGCTGATGACCAAGCAATTGCATTACCATGCACATCAGTAATCAT
 AACAAATAGTGTATTAAATGTAGCGTGAATATGAGCAATACCAAGATTCGATATTCTTT
 CACACGACGTTACGTGTTAGCCAAGACTTTACCTCCTATATTATTTTCTTA
 CCAGCAATCGCAACAGCTTACCTTACGAGTGCAGGGCGTTAGTGTGTTGTCCA
 CGGACAGGAAGTCCACGACGGTACGGATACCACGGTATGAACCGATTCCATCAAACGT
 TTGATGTTCAAGTTACTTCACGACGAAGGTACCTCAACTTGATTGCATCCACTTCA
 CGACGGATAGCATCTTCTTGATCTGATGTAAGATCACGTACAGAACATCTTCTGAGATT
 CCAGCAGCAGCCAAATTTCTTAGATGTCAGTCCGATACCATAAACATAAGTCAAT
 GAGATTACTACGCGTTGTCATTGGAATATCAACTCCAGCAATACGAGCCATGTTCT
 CCTTCTATCTTATCCTTGACGTTGTTGTTGGATTGCTGGCAAATTACCATAA
 CACGACCATTACGACGAATAACTTACAGTATTGCAAGTGGTTGACCGATGGTCTTA
 CTTTCATTCTTATCCCTCAAGTTTCGATTATTAAAGCGGTAAAGTGTACGTCCAC
 GTGTCAAGTCATATGGACTCATTGACAGTAACACGATCTCCGCTAAATACGAATAT
 AGTTTTACGAATTTACCAAGAAACTGTTGCTAAATCTGATGTTCAAGTTCCA
 CCGTAAACATTGCATTCCGGCATT

ORF Predictions:

ORF #	Start	End	Direction	Length
8	1147	1503	R	119 aa

[SEQ ID NO:] 3864344-8 ORF translation from 1147-1503,
 direction R
 VKKNIESGIAHIHATFNNTIVMITDVHGNIAIWSSAGALGFKGSRKCTPFAQMASEAAA
 KSAQEHLKSVEVTVKPGSGRESAIRALAAAGLEVTAIRDVTPVPHNGARPPKRRV*

Blastp and/or MPSearch Result:

Description:

30S RIBOSOMAL PROTEIN S11 (BS11). - BACILLUS SUBTILIS.

Assembly ID: 3864352

Assembly Length: 2627bp

[SEQ ID NO: 3864352 Strep Assembly -- Assembly
id#3864352

ATCGAATTATCTTGTATTCGCTGCAAATGGCTAGATGGTAAGAAGTAGACCGACTGAC
TAGCCTATAAACACCCGTTAAATCGCTAAGAACGTCAAAAAAGCCCTTAACATGGCAC
TAGTTAGGGCTTGGTCTAATGAACCTATACACTAACTACATTCTAGCATATAAG
CCCAGATATTCAAGAGTTATTTAGGTTCCCTAGTCTGAAAGGTCTAT
AATGAAGTTAGCCATCTAGTATCAAAAAACCGACTAGCTTATGAACACTAGTCGATTCT
CATCAATGCGCAACATTCTGAGCGATTCTGGCCAGATAGGTTATCTGGTAGTAG
GTTGCCAGTTGCCATTCTCAAAGAGGGCTTGGCTTGCTGCCTCAAAGAAGATA
TGGAAATGTTCTGCCTTAACGGGGCGATATTGTGGTCACTAAACTGAACATACTGAAT
TGTCCAGCGTCAGCATCTGTTCAAAGAGGAAACGCACGCCACGATTGCCCTTCTG
TAAGTCAAAATTCTTACCGACATACTGTAAGTGTATTCTGCTTGCTTGTCCACCTTGA
ACAAATTCCATAGTATTATCAGTAATGTTAATCTTAGTCACATCTGCTGATAGCCTTT
GTATAGTAAGCCTTGTACTCAGCCTGGTCATCTTACCAAGTCACCTAGCCTTGTAGTC
AAGACTTGGTCAAACGTGCCGTTCAAGGAAAGGATAAAACTGATTGCCAGTTACCTGCA
TAGTCACTCAAGGTGCCGTCCTGACAGCTGCATCCTCGAAGTAACCATTGGACTGTC
TTGGTATCCTCTGCCTTTCAGGTTCGATTGCTGGCCTTCTGGCTGTTGTTGTTTC
AAAGCCTTGAGGTTTCTCCATCAGGAAATGTAGTTCTCCAGCCTGGTGTCT
TCTGTCAGACTTCTAAAGGATTGAGGACATCAGTTGACACCTGCTTCTTGAAAGT
GTGTTAGCAAGGGCTTGTGAGGCATTCTCAAATAGATATAAGCGATTGTTATTTTC
TTGACATACTCTGTCATTCTGCCAAGCGAGCAGCTGATGGCTCTGCATCTGGAGAAAGG
CCTGAGATTGCGACTTGTGTTGAGTCATAGCCAAGGCAAGATAGTTAAAGGCTGCGTGT
TGAGTCACAAAGCTTTGTTGCTTGAGACAAGCCTCTGCGTAAGCCTTATCCAAG
GATTGCAATTTCGATATAGGCAGCTGCATTCTCTCAAAGGTCTCTTTTATCAGGA
TAATCTGCTGACAAGCTGCGGATGTGCTCTACTAGTTAATGGCACGAACTGGTGAT
AACCAAACATGGGGTCAAACATGGTGATGACCTTCTCCATGGTGTCT
TCTTCTCCTGCCACCTGGCAAGAGCAACATATCGCCTGCGCTTGTGGTTTCACT
TTTCTTATCCAAGGTATCTGCAATTAGGTACCCATGTTCCATGTTCAATT
TAAACGAAGGTATCTGCATCTGGATTTGGCAACTGCCTGGCAGATGGTGTATTCA
TGAGGTTCTGCCCAGCACCGATTAGGAGTTCTACATTAGCGTATCCTGCGACTTGC
TTGGTAAATTCTAGACAGGGTAAAGGTTGTCACGATATTGAGTTACCATCTGCCTGT
TTTGATTGAAACAAGCCACTAAAACAAGGCACATAGACTGGCTAGTAATAAGCTAATT
TTTTCACGTTCGTCTCCTATTGATAAAACGTCTTACTAAACTGATTAGTATAAAGACA
GTTACAAAATAATGGTAATACTTGCACTTGCAGGTGTTCTGCATAGTAGGAAATGTA
AGTCCTGCTACCATTCCAAAAAGCCAATCGCACTGGCAAGCAGCATAACCGATTAAAG
TTTTCCCCAGACGCAGGGCAACTAGCTGGCAAGACCATAATGGTCGATACCGAGAAGA
GCTCCTGCTGCAGGAATCATAAGGGCAATAGCCACCCGTACCATGTTAAAAGAATG
GACATGGTACGAAGTGGCAAGCCATCCACAAAGGCCGTATCTCGTCAAAGTTAAGATA
TACATAGGACGAAGAAAGAGAAAGGTCAAATCAAAACAACCGCCGCAATGACAAGAGG
GAAATGACCTGTTCTTCACTGATAGTCACGATCGAACCAAAGAGATATTGGTCAAAC
ATTGAACTCGAGTTTACCCCTGCTCATGACAATCAGAGAAACAGCCAGACCTGTTGAC
ACGAGGATAGCTGCCCCATTCCATAAGCTTTGTAAACCGTACGGAGATACTCCAGA
AAGACCGCCGCAATCAAGACAATGGCAATAGTAGAAATAGTTGGAGAAATCCCCAAAACC

AGACCNAAGGATACACCTGAAAATGAGACGTGGCTAAGGGTATCANTCATCAAACCTCTGA
 CGACGCACAGATGAGGAAGGTTCCAATACCGNTGAGTAAAGACTCATAGCAATAACCGC
 CAAAAAGGCCGCGTTGTATAAAGTCGTAAGATNATAAACTAAGCATGG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	303	1808	R	502 aa
7	1818	2528	R	237 aa

[SEQ ID NO:] 3864352-6 ORF translation from 303-1808, direction R
 VKKISLLLASLCALFLVACSNQKQADGKLNIVTTFYPVYEFKQVAGDTANVELLIGAGT
 EPHEYEPSAKAVAKIQDADTFVYENENMETWVPKLLDTLDKKVKTIKATGDMLLPGGE
 EEEGDHDHGEEGHHHEFDPHWLVLSPVRAIKLVEHIRDSLADYPDKKETFEKNAAYIEK
 LQSLDKAYAEGLSQAKQKSFVTQHAAFNYLALDYGLKQVAISGLSPDAEPSAARLAELTE
 YVKKNKIAYIYFEENASQALANTLSKEAGVKTVDVLNPESLTEEDTKAGENYISVMEKNL
 KALKQTTDQEGPAIEPEKAEDTKTVQNGYFEDAAVKDRTLSDYAGNWQSVYPFLEDGTFD
 QVFDYKAKLTGKMTQAEYKAYYTKGYQTDVTKINITDNTMEFVQGGQSKYTYKYVGKKI
 LTYKKGNRGVRFLFEATDADAGQFKYVQFSDHNIAPVKAEHFIFFGGTSQEALFEEMDN
 WPTYYPDNLSGQEIAQEMLAH*

Blastp and/or MPSearch Result:

Description:

ADHESIN B PRECURSOR (SALIVA-BINDING PROTEIN). -
 STREPTOCOCCUS SANGUIS.

[SEQ ID NO:] 3864352-7 ORF translation from 1818-2528, direction R
 VRRQSLMXDTLSHVSFSGVSXGLVLGISPTISTIAIVLIAAVFLEYLRTVYKSFMEIGTA
 ILVSTGLAVSLIVMSKGKNSSSMSLDQYLFSGSIVTISEEQVISLFVIAAVVLILTFLFLR
 PMYIILTFDEDTAFVDGLPVRTMSILFNMVTGVAIALMIPAAGALLVSTIMVLPASIALRL
 GKNFKSVMLLASAIGFLGMVAGLYISYYAETPASASITIIFVTVFILISLVRRIK*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3864366
Assembly Length: 1841bp

[SEQ ID NO:] 3864366 Strep Assembly -- Assembly
id#3864366
ATCGAATTCGAACTAACGATAAAAGGGACATTGAAAGCATCAACTTGCACATGGGACCC
TTTATCTTATGGAGGAGTTATCAGGATACAAAGAAATGGTCAAGATAACTTCTGG
TACCTTATTGACCATTGGCAGGTGAAGTGTCAAGTTGCACGTATCCAGTATCT
CTTGAGCTCAGAGATACAGAACGGTCAAAATATTGTATGGCATAAAGGGTGTGGA
AAATTCTCTAGAAAATCTTCATGCAATCGGAATGGATTAAAGTTGCCTTATCTTCATGA
GATTGATGCCTGCAAGGTTCTGACTGGGTTCAAATACGGTATGGTATCAGATATTCC
TGAAAGGTTGCCAATGGAATGCTCTATTAAACCCAGAACGGACTTTAGACTGGGATTC
ATCTGTCACACCTAACAGAGCGATGATTCTTGGTGGTATTACAGGGATTATTGATCA
TATGGATTACTTGCAAGACTTGGTATTACTGGACTATATCTTGTCCATCTTGAATC
TACAAGCAATCACAAGTACAATACGACAGATTACTTGAATTGACCGTCATTTGGAGA
CAAGGAGACCTTCGGAACTGGTGGATCAAGCGCATCATCGTGGCATGAAAGTCATGCT
GGATGCGGTATTAATCATATTGGTCGAATCTTCAATGGAAAAATGCGTAAAAA
TGGTGAACAGTCTGTTATAAGGATTGGTCCATATTCAACAATTCCCAGTGACAATG
AAAGCTAGTTAATAAGAGAGACTTACCCATCATGTTGGTTGGACTATATGCC
TAAGCTAAATACAGCCAATCCAGAGGTCAAGAATTATCTTAAAGGTTGCGACTTATTG
GGATTGAAGAGTTAATATCGATGCTTGGCTTGGATGTGGCTAATGAGATTGACCATC
AGTTCTGGAAGGATTTCGTAAGGCAGTTAGCTAAAATCCTGATCTTATATCCTAG
GAGAAGTCTGGCATACTCAGCCTGGCTAAATGGAGATGAGTTCCATGCCGTATGA
ATTATCCTTATCTGATAGTATCAAGGACTATTCCTACGAGGAATTAGAACAGAC
AGTTCATCGATGAAATCAATGGAGAGTTATGTATTACAAGCAGCAGATTTCAGAGGTCA
TGTAAATCTCTGGATTACATGATACAGAGCGAACCTGTGGACGGCCAATGAAGATG
TTCAACTGGTAAATCAGCCTTAGCCTTCTCTTTACAAAAGAACACCGTGCATT
ATTACGGAACCGAGCTAGCCTTGAATGGAGGACCAGATCCAGATTGTCGTTGTATGC
CTTGGGAACGTGTATCAAGTACAATGATGCTGAACCTATGAAGAGGCTGATTAAA
TTCGGAAATACCGTCAGTAATCATTCGATGGCAAGTATAGCCTCAAGAAATCAAAT
CTGATCTAGTAGCTCTGGAAATACGAAGGACGGATCCTCAAGCAATATTCAACC
AATCAACAGAACGATTATCTTTAGAGAAAGAACAGTAGCAAGCAATTGCCAAG
AATTGGAGAACGCTTGTCACTCTCCAGATGGATTGTGATTCTAAAATAGTTG
ATGAAGATTATGGTACATTTCATATCTTATAGTATAATAAGGCTAGTTACTAAACTTG
TAAAGGAGAACCTAAATGAATTGTAGAGGACATGAAACAAGACAAAGAACATTGTTAGAGAT
TTTGAAGTTAGCCTAAAGCACATATTAAGCTGTTAGCAAA

ORF Predictions:

ORF #	Start	End	Direction	Length
7	939	1670	F	244 aa

[SEQ ID NO: 3864366-7] 3864366-7 ORF translation from 939-1670, direction F

VANEIDHQFWKDFRKAVLAKNPDLYILGEVWHTSQPWLNQDEFHAVMNYPLSDSIKDYFL
 RGIKKTDQFIDEINGEFMYYKQQISEVMFNLLSDHTERILWTANEDVQLVKSALAFLL
 QKGTPCIYYGTELALTGGPDPCRRCMPWERVSSDNDMLNFMKRLIKIRKYASVIISHGK
 YSLQEIKSDLVALEWKYEGRILKAIFNQSTEDYLLEKEAVALASNCQELENQLVISPDGF
 VIF*

Blastp and/or MPSearch Result:

Description:
 neopullulanase (EC 3.2.1.135) - Bacillus sp.

Assembly ID: 3864384
 Assembly Length: 2026bp

[SEQ ID NO: 3864384] 3864384 Strep Assembly -- Assembly
 id#3864384

CTGTTAGCTGGTAAAGCCTTGATGAATTATTGACTTCGACGAATGTATTCCAGA
 ACCAGCAGCAATACGACGGCGACGGCTTGGATTAAACAAATCTGGGTTTCACGTTCTTC
 AGATGTCATCGAAGACACAATGGCACGTTACGAGCAATCTGGCGTTCATCCACCTTCAT
 GTTTGAAGTGCTGGATTGTTGGCCATACCTGGAAATCATCTTGAGCAAGTCTCCATCGG
 CCCCATATTTGCACCTGATCTAATTGATCGATGAAATCATTAAATCAAAGGTGTTTC
 GCGCATCTCTCAGCCATTCAAGGGCTTTGTCATCGTATTCCCTGAGAAGCTTCTC
 AATCAAAGTGAGCATATCCCCATGCCAAGGATAACGGCTAGACATAACGGTCTGGGTGGAA
 GGTTTCGATATCTGTAATTTCACCTGTACCAAGTGAACTTGATTGGTTCCAGTGAT
 GTGACGAACAGACAGAGCAGCACCACGAGTATCACCACATCAATCTGGTAAGGATGAC
 CCCAGTCACTTCCAAGTGGCATTAAACTCACGCCAACATTGGCTGCTCCTGACCAAT
 CATAGCATCAACGACAAGCAAGATTTCATTTGGTTGAGCCAATACTTCACATCACGAAG

CTCATTGAGGAGCTCATCAATCTGCAAACGACCCGCAGTATCAATCAAGACATAGTC
GTTATGATTAGTTGGGCTTGCTCAAACCTTGACGTACAATCTCAACAGCTGGTACTTC
TGGTCCAAGTGCAGAACAGACAGGCACATCAATCTGTTGCCAAGGTCTTAAGCTGGTCAAT
GGCAGCTGGACGATAAATATCCGCCAATCATCAAAGGACGAGCATTCTTCTTCTT
GAGTTGTTGCCAATTACAGCAAAGGTTGTTTACAGCCCCTGTAAACCAACCAT
CATGATGATGGTTGGAATCTTAGGTGACTTGATAATTGATCTGCCGTATCAGAACCTAA
AACGGCTGTCAGTTCTCATCAACGATTTAATAATCTGTCGCAGGATTAAGTGTATC
AATGACCTCATGCCCGACTGCACGCTACGAACCTTCTTGATAAAAGTCCTTACAACAGG
CAAGGCAACGTCGGCCTCGAGCAAGGCCAAGCGAATTCTTGGTTGCCCTTGACATC
AGATTTCAGAGATTTCTTTTACGTAGATTTAAAGACGTTCTGCAAACGTTCTGT
TAAACTTCAAATGCCATTCTCCTCTATTCTCTATTATCAATGCTGTTAAAATT
TCTATCTGCTCCTGCAGAAAATCATCCTGGGATAGCGATCCAAGATTGGTCAAAATC
TGACTACGGACAATGTAGTCCGAGTACATGTCAATTCTCATCTCATATTCTCAGCT
TTTCTGTTGCTTGATATTGTCATAGACAGCCTGACGACTAACACCAAACCTCAGCT
ATCTCAGCAAGACTGTAATCATCAGCGTAGAAAGCTCTATATAATTCTATTGCTTATCT
GTCAAAAGGCCGCATAAAATTCAAAGAGCGGCCATTCCATACGATTGGTTTTCGA
TTTCCATAACTTTATTATACCAAAAAATAGCCTAATCTACCACACTAGGGAGCCAATCC
TTGAAGATAGAAAGTAGATTGAGAAAAACGAGATCCTAGCCCCAAGTAATTCCAATTG
ATAGCTGGCAAAGGGATGCCCTCTGATTGTTGATAAGCTAGCTCAATTCTTGC
CCTATCAACTTGATAACGGCTGTTGAATGATAAAATTCTGATGCCCATAGGGTAGG
AATATAGGCCAAACTATCACTATCCTTAAAAAGCGCATAATGGTCTGGGATTAGAAAA
TCGGCTCATCACCAGTTCTGACCAGTAAATAACTACTTTCTTCTTCATT
ATGAAAGAGTAAATAGCTATAATCTCCCTTCTGACTTCCACAA

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
8	1717	2025	R	103 aa

[SEQ ID NO: 3864384-8] 3864384-8 ORF translation from 1717-2025, direction R
VEVHEKGDSYLLFHNEEKEKVVIFHGQELVMSRFSNPKTIMRFLKDSDSLAYIPTPMG
MQEFLIIQTSRYQVDROKIELAYOLONOEGHPFASYOLEITWG*

Blastp and/or MPSearch Result:

Description: unknown

Assembly ID: 3864400
 Assembly Length: 1561bp

[SEQ ID NO:] 3864400 Strep Assembly -- Assembly
 id#3864400
 CTTGATTATGGCTGTTGGAAAAACGGGCAGGGCTCTCTGCAAAATCAGGATGCCTA
 TCTCAAATCTGCTGGTGGTAAATTGGATGAACCTGCCATTGACTGGCTGGCAGT
 TGCTATTGCTCGAGCTACAAAGACAAGCCAACTAATCCTCAGGAATGTTTGTGGAGA
 ACTGGGCTTGACAGGGAGAGATTGGCGCGTGAATCGTATTGAGCAACGCATCACGAAGC
 TGCTAAACTGGGCTTACTAAGATTAAAGTACCTAAGAATTCTTGACAGGAATCACTCT
 GCCTAAGGAAATTCAAGGTATTGGCGTGACAACGATTCAAGGAAAGTTGAAAAAGGTCTT
 TGCATAATCCGTGACAAATTCTCTTAAAGATAAGATAGGGAGAAATATTTGACTATCA
 AATTTCAAGGAGGGAACTCGTGTGTTGAACAGTTATGCAAGCTAATCAGGCTTA
 TGTTGCCCTACATGGGCAGTTAAATCTGCCACTTAAACCCAAAACAAGAGTAGCTATTGT
 GACCTGTATGGACTCTCGTCTGCACGTTGCGCAAGCTCTGGCTTGGCACTTGGGATGC
 TCATATCTGCGGAATGCAGGTGGTCAGTGACTGAAGACATGATTGCTCGCTAGTTAT
 TTCCCAGCAACAAATGGGACAAGAGAGATTGTGGTATTGCACCATACAGACTGTGGTGC
 TCAGACCTTGAAAATGAACCTTTCAAGGAGTTAAAGAGGAATTAGGTGTAGATGT
 GTCAGACCAGGACTCTGCCCTTCAAGATATAGAAGAGAGTGACGCGAGGATATGCA
 ACTGCTTATCGAGTCTCCCTAATACCAAGACGATGTCATTATCTCTGGTGTATTTACAA
 TGTTGATACAGGAAGTATGACAGTCGTAGAATTATAACTTCATTAGAAAGAAAGTG
 TATGAAGAAAAGCAGTATTTATTGCTATGTATTGGTTACAGTATGAAACCACATCTACTA
 TACGGACGGTCCAAGGTGCGGAATATGGACTAATGGAGTTCTATCTTAGC
 TCTCTTTACATGATTCCGGCTTTATTCTCTCCATATTGGAAAAAAATGGGAATT
 GCCAAAGAAGGTTTGATTCTGCTTTATTGGGAGCAATCTGTTCTTACTCTCTCTT
 ACTATTGGAATCTATAATCACAGACGAAAGTCATCTAAGGTATAAAAATCGACCAGTT
 ACTGGGGTTCTTCCCAGATAGTACATTAAATGCCATTGAAAGTGCTATTGTGGC
 TCCTTGGTAGAAGAACCTTGAAATTGCACTTGTGTTTGGCTTGGGATTCAA
 ATGATTGAGGATATTGGTTATTCGTACGGATTGCCAGAGGGCTTGACTTACTATT
 TCGCGAATTAGAGCGTATCATCTCAGGAATTGCCTCTCACTGGACTTTTCAGGTCTA
 G

ORF Predictions:

ORF #	Start	End	Direction	Length
7	371	937	F	189 aa
			103	

[SEQ ID NO:] 3864400-7 ORF translation from 371-937,
direction F
VTNSLKNDKIGEIFDYQIFKEGIVSYFEQFMQANQAYVALHGQLNPLPKTRVAIVTCM
DSRLHVAQALGLGDAHILRNAGGRVTEDMIIRSLVISQQQMGTRIEIVVLHHTDCGAQTF
ENEPEFQEYLKEELGVDVSDQDFLPFQDIEESVREDMQLLIESPLIPDDVIISGAIYNVDT
GSMTVVEL*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3864416
Assembly Length: 2009bp

[SEQ ID NO:] 3864416 Strep Assembly -- Assembly
id#3864416
AATGATTTCAAGCAGACGATCCATGTCATTCAAGGAATACATGCGACGATTCCCTTC
GTTTCGATCGGGCTTGTCAACTCTTGATCTTCATAATAACGAATCTGACGCCGATAG
ATCGGTCAACTCATAAACACTGCCGATAGGAAAAACAGCCATATTCGGCGAAATTCTTT
TTCCTTCATTACAATTTCCTTCTTGTCTATTATAGTCTAAAAAAAGACAAACGTCA
ATTGATAATGTTATAAAATGTAACATTATTTTCTTATTCTCTAAAAAGAGACGAATAC
GATCAATATCGTAATTACGATAATTGCGACAAAAACTCCCATAAACGTTCTAAAACAC
GCACAAACACGTACAAAATTGTCACCACCTGGAATTGATAGGGTAATGATTAACATAG
CTGCTACACCACCAATAACCCCTGTTGTTATTGATGGCTACATTGTCATAATGGTTA
ACATGGTGCAGATTGGAACAACCTACCAAGGTACCCAAAAGGCTCGTGGAAAAGGTAT
TTAATAAGAAGAAGACCAAGGCATAGAGTCCACCGATACTATTCCTAGAATACGCGAAG
TCCCAAAATGAACACTCTCATAAAACCTCCCTCAGGCTAAAAACGGCTGTCAAAGCAC
CAATTGAAAGACCTTCCAGCCAAAAAGCCAAAATCAAGAGAACTAGAAAACAGCAA
TACCTGTTAAAGGTTGCATACCAAGTTGAACGGGATTATCGAATTATTTT
TAAAATAACTCATAACTCAACTTCTATTCCATTTATCATAATCGGTATT
GAGTAATAGTTGAGAGGAAGCGTTTATTAAAGCAAAAGAAAAGAGGAACCTTCATCC
CTCTCTTCTTGATTATTATAAAATCTTATTCTGTCAAGGCTGCAAGTCCTGGAA
GAACCTTACCTCAAGAAGTCCATTGATGCTCCACCACCGTACTAATCCATGAGAACT
TGTCTGCACGGCCAAGGTTAACGCTGCGGCAGCTGAGTCACCACCGATGATTGATT
TAACTCCTGGTTGTTCACGATAGCGTCCATCACACCGATTGTACCAGCTCTGGAAATCT

GGGTTTCAAATACACCCATAGGTCCGTTCCATACAACGTAGTTAGCACCAGTCAGCTTCAAGCT
 TCGCAAATTGGCGATAGATTGGACCGATGTCAGACCAAGGAAGCCTCAGAAACT
 GCTTCACCTTCAGTGTCAACGCACCTCAGTGTAAACCAGCAAATGCAGCTTAGCTCTTTGA
 GTCAACTGGCAAGATCAATTACCATTTGCTTTCAAGAAGAGCTTCGCAACATCCAG
 TTTGTCTTCTTACAAGTGAGTTACCGATTGATACCTGTGCTTGTAGAAATGTGTA
 AGTCATCCCACCACCGATAAGGACTTATCAGCTTTCAAGCAAGTTCGATAACACC
 GATCTTGTCTGAAACTTTGAACCACCAAGGATAGCCACAAATGGACGTTCTGGAGTTTC
 AACTGCTTCTTGGATGTAGGCAATTGCTTTCAAGAAGGAAACCAGCAACTGCTTTTC
 AACGTTGCTGAGATACCAACGTTAGATGCGTGTGCACGGTGAGCTGTACCGAATGCATC
 GTTACGAAGATACCATCTCCAAGTGTAGGCCAGTATTACCAAGTCAGGATCGTTTT
 AGATTCTTCTGCCGTCAACATCTCGTAACGAGTGTTCAACCAAGAGAACTTGTCC
 ATCTTCAAGAGCGTTGATTGCCGCTTCCAATTCAAGCACCACGAGTGACACCTGGAAAAC
 AACATCTTGACCAAGTTGCTGCCAAGTCAGCTGCTACAGGAGCAAGTGATTACAGC
 TTTATCAGCTTCTTCTTCAACACGTCCAAGGTGAGAGAAAAGAATTGATGTCCACCTTG
 TTCGATGATGTAAGTTGGAAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	929	1189	R	87 aa

[SEQ ID NO: 3864416-7] ORF translation from 929-1189,
 direction R

VLKQLYGTDLWVYLKTQISRAGTIGVMDAIVKQPGVKSIIIGGGDSAAA
 INLGRADKF SW
 ISTGGGASMELLEGKVLPGGLAALTEK*

Blastp and/or MPSearch Result:

Description:

PHOSPHOGLYCERATE KINASE (EC 2.7.2.3). - YARROWIA LIPOLYTICA
 (CANDIDA LIPOLYTICA).

Assembly ID: 3864424
 Assembly Length: 2299bp

[SEQ ID NO: 3864424] 3864424 Strep Assembly -- Assembly
id#3864424

TGTGAAAGAGTCCATGGTCCGATGGCAGCGTTGGGTAGGTCTGCCAACTGGCGACCCAA
GTGTTGTTGAGCTGACATCATCTGTTCTGGATTTCTTGCTGATTTCCTCTCA
AACGTTCTTAAGTTAGTCAGTGACGCCAGCGTAAAGGTTGAGATAAAAGAGTTGAGAAA
TTTCGACACCACGCCAATTGGTCCAGAATGCGCTCTGCCATGACAAAGGTCTTCAG
AACCAGCCGATGCTGAGACCAGGATATTCTGCCAGAAGTGTAGATAGCTTCGATTGCT
CGGCAGTTCTTGCTCGAATTGCTCTGCTTGCAGTTGAATCT
CCTCCTCACTTAAAAGGAAATAAGCTTCATCGATTCAACTCCTCTCAATTTCAC
CCAAGCTTGCTTGAGTTCTCCGACCAGACGCTTGCTATCAGCTAGGTCAAAC
TAGGAAACGGGCTTGGCCCAGATGGTAATTGGCTCAAAGCCTGTAATAGCCTGATGTTG
CTGGACGTATGGGGCAATGCTCTGCCATTTCAGTATAAGGATTGATGGCGAACCGGCC
TGCTAAAATCTCTCAGCAGCTTCTGTAAAGATAGGCATTGTAGTCCAGTAGGAGCTG
AAATTCCATCTGTCAGTTGATTAGCCTGTTTTGTTAAAATCGCCTAAATAACT
GCTTCTTTCCAAGAAGAGCCCTGGTATTCATAGATTGCTGGCTCTACACTGC
TCCTGCAAGACTTTTACCGCCATCAGAGATTGGACAGGTTAGCCATTCAAAGTACAT
GGCGCCGAAAAAGTTCTGCTCCCTCTTTAGGGCAGCAAGATAGGTTGGTAAC
AGAATTGAGCCCATTAAAGAAATGAGGAAACTGGAACGTAGACTGGATTGAGTC
TACTACTCCTATCGCTCCATTAGCTTCAAACGGTCAATCCGGTCCACCTGCGTAC
AAAGACACTGCGTCCATTGTCTAATTGAATAAAGGCTGGCTTTCCACAAAATTG
TTCTCTTTGATGGTTTCGATGGCTGGATTGTGCGAGAATATGTCAGTCCAGTGT
AACATCAAGCAAAACTCCTGGTAAACTGGCTCCAAACTTCTGATAAAATAGCTC
AAATTCGCGTTCTGACTGGTTCTGAATAGCTTCTAGACGTTGGTCAAAGGAATC
TTCATTAGGCAACTGTAAGGCGTTCAAAGATACTGCAAGAAATTCCCGTACTACG
GGCATCAGGATGCAAACGAATTCCCTCCTGCAAGCCTAAACGTAGCGTAGGAAATAACTG
TATTCATTGCGATAAAACTCTGTCAAACCCGACGTAGACAGGTAACACTCCTGTTGG
GGATAGAGAGCTGCAAGGTGCTGGCTAAGGTCTGCTGCTGGACTGATTGGATG
GCTGGATTTCCAGACCTGCTGATCTAGTTTACCTATGACACCGCAGAACCTTG
ACAAAAGTCAAATCTGCTCAGTATCGCTCATCTCACCCCTGCTGGTATAGGCAACCAGA
CTAGACAAAAGACTGTGATAGGACCCCATATCCTCCTAGACAGTCCTTGTGATT
CTCTCTCTCCGCCCTAAATCCAAAATGGATCAACTCTGAAAGATAGGCAGATT
CTTCACTTCGTTAAAAGGCTTGGAGCCGACAAGAACACTGCTACGAGCAGAATTG
ACCAAGGAAAGCATAGTGTAGCGATTTCTGAGATTTCACTGCTGGCAATCAGTA
TGAACGCCCTTCGGCGCTGGTTAGGTTTGCTTCTTCATCTGTCAGAAGACTG
GTGTTTGAGAAATTGGTAAATTGATCCTGAGTTAGTCCAATAGCAGACAAAGT
CAGCAGTCAATGGTCAATCAAATCGTAACCTGCACCCAGAACAGTGTCCACTG
GAATGGTACGGTATTGGGACAAACTCATTCCAGAATGGAGCAAGGCTAGGAAGT
GACTAACCTGTGAACCAGCAAAACAGTCGCAAATTGTTCTAAACATGGCAGAAAGC
TCCAAACTTCGGCTTGTCTTCTGTTCTACAGCTTCAAAGTGGTGTCAAATCTG
ACTGCTTGGTCACAGCTCCTCTTTAGAAAGACACTCCATTGTTGAGGAGTT
CCTTTGTTCCGCTGGC

ORF Predictions:

ORF #	Start	End	Direction	Length
7	388	1008	R	207 aa

[SEQ ID NO:] 3864424-7 ORF translation from 388-1008, direction R

VDRIDRLKANGAIGVVVDYKSSLTQFQFPFFNGLNSQLPTYLAALKREGEQNFFGAMYLE
 MAEPVQSLMAVKSLAGAVVEASKSMKYQGLFLEKESSYLGEOFYNKNKANQLTDEEFQLLL
 DYNAYLYKKAEEKILAGRFAINPYTENGRSIAPYVQQHQAITGFEANYHLGQARFLKKLD
 LADSKRLVGEKLKQAWVEKIREELNR*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864430

Assembly Length: 1915bp

[SEQ ID NO:] 3864430 Strep Assembly -- Assembly
 id#3864430

AGAGGTAGTCGTAAACGTAAAAAATTCTAATTGAAATGAAAGGGCTAGAGGAAATCTAG
 TCCTTTCTTTAAATAATACTCCAAAGCCTGCAAAATCTGAAACTTCCTCCTACAA
 TTTGATATAATAGAGAGAAGAATTCAATTGAAGGAGGAAATGATGTCGGTTTAGTAAA
 GAAGTGATTGAAAAGCTTAGACTAGATATTGTCTATGGTGAACCAGAAATTGCTGAAAAG
 GAAATCAATACAGCGGATATTACCGCACCTGGTCTGAAATGACAGGCTATTTGACTAC
 TATACACCAGAGCGGATTCAACTTTGGGATGAAGGAGTGGTCTTATCTGATCAGCATG
 CCTTCCAACAGCGTTATGAAGTTTGAAAAAAATGTTCTACCTGAGACACCCAGCAGTC
 ATTGTTGCCGTGGTTGGTCCAGAGGAGATGTTAAAGGCTGCTAGAGAATGTAAG
 ATTGCTATTTAACAGCCGTGCAGCTACCACTCGTTATCTGGAGAGTTATCTAGCTAT
 CTGGATTCTCGTTGGCAGAACGTACCACTCGTGTGACCGTGTCTGATGGATATTTATGGG
 ATGGCGTCTGATTTCAGGGAGATAGTGGGAATTGTTAAGAGCGAGACAGGTCTTGAGC
 TTGTCAAACGTGGTCACCGTTGGTAGCCGATGACCGTGTGATATCTTGCCAAGGATG
 AGATTACTCTGGGGTGAACCAGCTGAAATTGAAACACTGATTGAAATTGCTGTCAAGGATT
 TTGGTATTATCGATGTTATGAGTCTACGGTGCGAGTGCTGTCAAGGATTCTCACAGG

TTCAGCTTGCTGTCTATTGGAAAATTACGATAACGCATAAGACCTTGATCGTCTGGAA
 ACAATGCAGAGGAACCTGAAGTTCTGGCGTAGCCATTCCCTGTATTCGTATCCAGTTA
 AACACAGGTCTGTAATATCTCTGTTGTGATTGAGGCAGCTGCCATGAATTATCGTGCCAAGG
 AAATGGGCTTGATGCTACCCGTTGTCGACGAACGACTGACAAGTCTCATAGCTCGAA
 ATGAGGTGCAAAATGCTGATCCAATTGCTATTCAACTAGGACCCCTAGCCATTGTTGG
 TATGCCTTATGTATTGTGACAGGCTTGATTCTGGGTTATTGACCATGAAAGAAGCA
 CCTAGAAAGAAGATCATACCAGACGATATTTAGATTTATCTTAGTAGCCTTCCCTG
 GCTATTTAGGAGCTCGTCTACTATGTTATTTCCGATTTGATTACTATAGTCAGAAT
 TTAGGAGAGATTTGCCATTGGAATGGTGGTTGCCATTACGGTGGTTGATAACT
 GGGGCTTGTGCTCTATATCTTGCTGACCGTAAACTCATCAATACTGGGATTTCTA
 GATATTGCGCGCCTAGCGTTATGATTGCTCAAAGTTGGGGCGTTGGGTAATTCTTT
 AACCAAGAAGCTTATGGTGCAACAGTGGATAATCTGGATTATCTACCTGGCTTATCCGT
 GACCAAGATGTATATTGAGGGGAGCTACCGTCAACCGACTTCCTTATGAGTCTCTATGG
 AATCTGCTTGGCTTGCGCTTGATTCTGATTTAGACGGAAATGGAAGAGTCTCAGACGA
 GGTCAATACACGGCCTTACTTGATTGGTATGGTTCGCTGAGTGTCCCAATGGCTGTCAGTTGTC
 ATGCGAACAGATAGTCTCATGTTCTCGGCCTCGAGTGTCCCAATGGCTGTCAGTTGTC
 TTTATCGGTCTCGGTATAATGATCGTTATTTATCAAATCGAAAGAAGGCCCCTACTAT
 ATTACAGAGGAGGAAAACAAATGTTAGAAGTTGCATATATTCTTGTGCCCTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	627	1100	F	158 aa

[SEQ ID NO: 3864430-7] 3864430-7 ORF translation from 627-1100,
 direction F
 VGIGKSETGLELVKRKGHLVADDRVDIFAKDEITLWGEPAEILKHLIEIRGVGIIDVMSL
 YGASAVKDSSQVQLAVYLENYDTHKTFDRLGNNAEELEVSGVAIPRIRIPVKTGRNISVV
 IEAAAMNYRAKEMGFDATRLFDERLTSIARNEVQNA*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864442

Assembly Length: 2245bp

[SEQ ID NO: 3864442 Strep Assembly -- Assembly
id#3864442
ATCGAATTGAAGTGGTTGAAGAGAGTACAACCTGTCTTTAGAAAAGGAGCCTATAAT
GAAAGTCTTCAGCATGTAAATATCGTACCTGTGATCAAGATTCCATGTTATCTTGA
TGGAAATCTTAGCAGTCAGGATTCTCAAATCGTCTATGTCGGTCAAGATAAGCCANCCT
TTTAGAACAGCTGAGCAGATTATAGACTATCAGGGAGCTGGATTATGCCTGGTTGGT
CAATTGTACACCCATTCTGCAATGACAGGTCTGAGAGGGATCCGAGATGACAGCAATCT
CCATGAATGGCTCAATGACTATATCTGCCAGCAGAATCTGAGTTACTCCCACATGAC
TACCAATGCGGTCAAAGAACGCCAACAGAGATGCTCCAGTCAGGAACAACACCTTAA
CGATATGTATAATCCAATGGTGTGGATATCCAGCAAATTATCAGGTGGTAAAACCTC
CAAGATGCGTTGTTATTTCTCCACTCTCTTCAGAGACAGAACAAACTGCTGA
GACTATAAGCAGAACACTCGATCCATCATAGACGAAATCTTAAATATAAAATCCAAATT
CAAGGTTATGGTAGCACCTCATTCTCGTATAGCTGCAGTAGAGACTTGCTGGAAGCGAG
TTTGGAAATGGCAAAGAGCTAAATATTCCGCTCCATGTCCATGTGGCGGAGACCAAGGA
AGAGTCAGGAATTATCCTCAAACGGTACGGCAAACGCCCTTGCTTTCTGGAAGAACT
GGGTTATTAAAGATCATCCGTCGTATTGCTTCACGGGTCGAATTAAACGAGAGAGAA
ATTGAACTTCTGGCATCTTCTCAAGTGGCTATGCCAACATCCTATCAGTAACCTCA
AACTGGCATCAGGAATTGCTCCAATTATCCAGCTCCAAAAGCGGGAGTAGTAGTCGGA
TTGCGACTCGGTTGCTCCAATAACAATCTAGATATGTTGAGGAAGGAAGGACTG
CAGCTCTCTTCAGAAGATGAAAAGTGGGGATGCCAGCCAGTTCCAATCGAAACAGCTC
TCAAGGTACTGACAATCGAAGGGCTAAGGTCTTGAATGGAAAATCAGATAGGAAGTC
TGGAAAGTCGCAAGCAAGCAGATTCTGGTCATTCAACCACAAGGGAAAATCATCTCC
AACCTCAGGAAAATATGCTGTCACCTGGTTATGCACTTAAATCTAGTGATGTAGATG
ATGTTTATATGCCGGAGAACAGGTTGTTAAGCAAGGTCAAGTCCTGACAGTAGAACTTT
AAAAGAAAATCACGAAAATTTAAAAAAAGTTCTGCAACAAATCTGCATTCTTTTT
TGACTATGCTACTTATACGGTTAAAAAAACTGCCTAAGACAGTAGGGGAGCTCGA
CTCATAAATATCCTACCGAGGACAAACGTATCATGTTAAAGAAGCGTATTGACTTTC
GTGTCTAGGTTGGCGCGTTTCTTTGAAAAATTCCCCAAGCAAAATAATTACGGA
GGTGAACACACTAATGAGTGAAGCAATTATTGCTAAAAAAAGCGGAACTAGTTGACGTAGT
AGCTAAAAAAATGAAAGCTGCTGCATCTATCGTCGTTGAGACGCTCGTGGTTGACAGT
TGAGCAAGATACTGTTCTCGTCGTGAGCTGGAAAGCGAAGTTGAGTATAAGTTAT
AAAAAACTCAATCTCGTCGTGAGCTGAAAAAGCTGGCTTGAAGATCTGCATCTGT
ATTGTTGGACCATCTGCAGTAGCATTTCTAATGAAGATGTTATCGCACCGCGAAAAT
CTTGAACGACTTTCTAAAAACGCTGAAGCACTTGAAATTAAAGGTGGTGCAATCGAAGG
CGCTGTCGACATCTAAAGAACAGAGATTCTGCACTTGCAACTCTCCAAACCGCGAAGGACT
TCTTCTATGCTCCTTCTGACTTCAGCGCCAGTGCAGCAACGTTGCTTGCAGTC
AGCGGTTGCAGAAAGCAAAGAACGCGGCTTAATCTTAAGCTACACAGCGTAGCCTAGC
TACGAAAAAAACTATTATAAAATTAAAGCTTATTGGAGGAAATAACAATGGCATTGAA
CATGAAAACATTATTGCTGAAATTAAAGAAGCTCAATCCTGAATTGAACGACCTTGT
AAAAGCTATCGAAGAAGAATTGCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
7	867	1322	F	152 aa
8	1562	2074	F	171 aa

[SEQ ID NO:] 3864442-7 ORF translation from 867-1322, direction F
 VAIANPISNLKLASGIAPIIQLQKAGVVVGIAATDSVASNNNLDMFEEGRTAALLQKMKSGDASQFPIETALKVLTIEGAKVLGMENQIGSLEVGKQADFLVIQPQGKIHLPQENMLSHLVYALKSSDVDDVYIAGEQVVVKQGQVLTVEL*

Blastp and/or MPSearch Result:

Description:

N-ethylammeline chlorohydrolase [Rhodococcus corallinus]

[SEQ ID NO:] 3864442-8 ORF translation from 1562-2074, direction F
 VNTLMSEAIIAKKAEELVDVVAEKMKAAASIVVVDARGLTVEQDTVLRRELRGSEVEYKVIKNSILRRAAEKAGLEDLASVFVGPSAVAFSNEDVIAPAKILNDFSKNAEALEIKGGAIEGAVASKEEILALATLPNREGLLSMLLSVLQAPVRNVALAVKAVAESKEDAA*

Blastp and/or MPSearch Result:

Description:

50S RIBOSOMAL PROTEIN L10 (BL5). - BACILLUS SUBTILIS.
 (BLAST)

Assembly ID: 3864450
 Assembly Length: 1471bp

[SEQ ID NO: 3864450] 3864450 Strep Assembly -- Assembly
 id#3864450

GGGAGAGAACTGTGACAGAAAAACCAACAAACTCGTTCTCTAAGTCAGAAGATTTGG
 TGAAGATTCCAAAGGGGAATTGCATTAGAAAATGATTGATGATGAATCTTCTATG
 GTGAAAAAGCTCTTGATTGGAAGGGGATGATTACAGGATGGCATAAAACAAAGATG
 GTAAGGATTATCTAGGATATAACAGTCATCCCTGCTAGCAGACAGTGTAGGGATGGTT
 TGGCAGATGGGAAGATGATAATAAGAAAGAATGGTATGTCACAGACCGTGATTCTCTC
 TCTTATGGAGTTAGCTTATCGAGACGATGATTATATTGAGAAAATTTAGATCATAAGA
 ATCTTTCCCTAGTCTCTATCTGACCGTCAAGAACACAAACTCATGCACAATGAATTGG
 CTCCTTCTGGAAGATGAAAAAGCCTACTATACAGATAGTGGCTGGATGCTTCTTAT
 TTGAGACCAAGAGCGACCTCCTTATCTCAAAGATGGAACGGTGACATGTTGGCTATT
 GTGGAACGCGAGTTAATGACGCCAAGGACTTGAGTGCAGATTGTTATTAGGTGGAA
 ATAAACTAGCTCAAGCGGATGATATCCGCAAGGTTGTTGGGAATTAGCCAAGGATATAA
 GTATTACTAAGTTGTATATGACAGGTCACTCTGGAGGCTACCTAGCTCAGATTGCAG
 CGGTTGAAGATTACCAAAATATCCTGATTTATAACCAGTATTGAGGAAAGTGCACAA
 CTTTCAGTGCCTCAAAGTCATTACTCCAGAACTGTTGGATGCTAAGAACATGGTTCT
 GAGATGTTGGTTGGAAAGTCGTAATTAGCTGTTAGTGGAAAATTAAGCATTATGTGG
 TTGATAATGACAATGTTGTGACTCCCTGATTCTATAATAATCGTGTATTGTTACATT
 CAGGTAATTACGCTTAAACACCGTTCTCGTGGCTATTGAAAGTCCAATGAATGATA
 TTCCTAACTTAATATTGGTAAACAAGCTACCTGGATAAACATGGTTATCGTGTACCGA
 AATTGGATAAAAGTGCATTCTTAAGAACAGGGCTCTACCTCAATCTCTAGTCACCAA
 GCGCTGAACCAATGGAAAATATGCGCTTAGGAAACAGGTACTCAAAGTCGACAGCCT
 TCGGAGGAGATGCTAGAAGAGCTGTGGATGGCAAAGTCGATGGTAACATGGTCACAATT
 CTGTCACTCATACAAACTTCAATCTAACGCTTGGCAAGTAGATTGGCTAAAGAAG
 AAACCATTGCCAAATCAATATTACAAACCGAACAGACACTGCCAGGATAGATTGGCAA
 ACTTGATGTCATTCTTTAGACAGTTCTGGTAAAGAAATCGAGTAAAACGTATAATA
 TCTCCTAAAGATGTCAGCACAAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
7	897	1448	F	184 aa

[SEQ ID NO: 3864450-7] 3864450-7 ORF translation from 897-1448,
 direction F

VVDNDNVVTPLIHNNRDIVTFTGNSRFKRSRGYFESPMNDIPNFNIGKQATLDKHGYRD
 PKLDKVRFFKKQALPQSSSQPSAEPMENIALGKQVTQSSTAFGGDARRAVDGKVDGNYGH
 NSVTHTNFQSKPWQVDLAKEETIRQINIYNRTDTAQDRLANFDVILLDSSGKEIRVKTY
 NIS*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3864482

Assembly Length: 1954bp

[SEQ ID NO:] 3864482 Strep Assembly -- Assembly
id#3864482
CTACGATAAAAGTCACCAAGAGTCATTAGCAGGTGCTGAACAAGTCCTCAGTTTTCTG
AAGCTTGGTCAAAAGTTCGATAACTTGGTCTGCAGATGTTGCTTGACGAAGTTGTCTG
CAAAACCGTCTTCATCAAGTATTGAGACAATTCTGTCAATGCTGCCAAGTGAGTATCAT
TGGCACCTCTGGAGCTGCAATCATGAAGAAGAGGTCAGTTGCCTGCCATCCAAACTCT
CATAGTCAACACCCTGTTGACTTAGCAAAGAGAACTGTCGCTTCTTGACAGCAGCGT
TTTGCTGTGAGGCATAGCGATTCCATACCCAAACAGTAGAAGTAAAGCTTCACGCG
CCAAAATGCCTCTTAAAGGTTCAAATCTGTACATAACCGTGCCTGTTAGGCTTT
TAATCATCTCTTCAATGACAGCAGTCTTCAGTTGCCTGCCAATCCAGAACATGACAT
CTTTCTCAATAAATCTGAATTTCATCGTTTCTACCTCAACTTTCCATATGTTCT
TTAATAAAATTCCGCCGTTGCCAAGTCATCTGAGAAGGGTAGTTGCCGTTCCGCAAGGCC
TCCCCATTTGAAGGCTTCACTGCGTCTTTGATTGACAAATTCAACCTGTGAATCCAGC
AACCATAGAACATCACCAAGCTCCAAGTGAATTTGACTGTTCTTGATTGGTTAGCGAA
GTAAGCTCCCTCAGATGTGACAAGAAGGGCACCACCAAGCCATAGAGATAATAACATT
TTGAGCACCCCTAGCCAGTAACTCACGAGCGTATTCTCAATTCTAAACTTCAG
TTAACCCCCAAAATCGCTCCAAGTTCATGATTATTGGTTTACAAGAACAGGCTGGTA
ATCCAAACTATCAATTAAAGGTCTGCTCCTCAAAGTCACAGACCACCTGCGCACAGCTG
GCGCGTCAAGGAAATCAAATCCTTATAGATAACATTGCCTAGATTGACTTGAACC
TGCAAAGACAACGTATCTCTGCTGTCAAGACTAGATAAAATAGCTTCAATTCTCTAG
CTTAACCGGTTCAACAGTTGGACCCGTTCCGTTGATTCTGTTCTGGTCTGCTTNGAT
TTAACATTGATACGAGTATCTCTGCCACCTGGACAAAAGGGTCTCGATTCTTCCCTC
TGGCTAAAGTATCTGTGATAAAATTACCAAGTAAAGCCACCGATAAAATCCCGTCTGTA
TTGATATATTCAAACGTTCAAGACACGGCTGACATTGATTCTTCCCACCAAGCAAAC
TTATCATCACTGTCCATACGATTACACTACCAACTTGACTTGGTCCAAACGAACGATA
TAGTCATGGATGGATTGAGTGTGACTGTATAAATCATACCTCTATTACCTCCGTTTCT
CCTTAATAACCTGCAAGAGCTCATGCCCTGACTAGTGATAACGATAGCGCGTTGAGTG
GGGCTACCTTGGCAAAGCAAGTTGTCCAATTGGACGAATCCACCAAGACGTAGGTCT
GCTTGGCATTCTCCAAAATAGCTCTTACAGCTCCCTCCATATCAGGAGTCGTAT
AATAGCCATCGTCAACACCATTCAATTCCGATAAAGGCACGGTCAAAGTGCAATTGGTTAA

TCTGGTTAAGAGCAACGCCCGATACTAGCATCTGTCGCCGTCTGACGTTCCCTCAA
 CCATGACAGTTGAATCTGCTTTCAACCAACTGAGCGGCATGGTGAATGGAGTTGGTCA
 CAACTGTAACATTCTTATTGACCAATTGATGAAATCAAAAAGCAGTTGTTGTTCCCAGCA
 TCCGATAAAGATGACATCTTTCCTTAATGAGAGAGGCTGCTTCTGAGCCAGCAATT
 TCTTTCTTGAAGGTTTGACAGATTTTCTTG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	505	1170	R	222 aa

[SEQ ID NO: 3864482-6] 3864482-6 ORF translation from 505-1170, direction R
 VAEDTRINVKIXADQETEINGTGPTEPVKLEELKAILSSLTAEDTVVFAGSSAKNLGNV
 IYKDLISLRTQTGAQVVCDFEGQLIDSQPLLVKPNHNLGAIFGVKLESLDEIEKY
 ARELLAKGAQNVISMAGDGALLVTSEGAYFAKPIKGTVKNSVGAGDSMVAGFTGEFVKS
 KDAVEAFKVGACGTATTFSDDLATAEFIGKETYGKVEVEKR*

Blastp and/or MPSearch Result:

Description:

1-PHOSPHOFRUCTOKINASE (EC 2.7.1.56) (FRUCTOSE 1-PHOSPHATE KINASE). - RHODOBACTE R CAPSULATUS (RHODOPSEUDOMONAS CAPSULATA).

Assembly ID: 3864496

Assembly Length: 1975bp

[SEQ ID NO: 3864496] 3864496 Strep Assembly -- Assembly id#3864496
 TCAAAGAGTAACAAAGGCACCAAATTCTCGATAGGAACGATTAGCACGGTAAACTTCAT
 CCACTTGGGTTCACGGAACCAAACCGAGCAATAATTCTTGGGCACGGGTTAATAGCAT
 TTGGTCAACTAGGAGTAGATAGAACACATTCNTTCTCGTCTATATCAATCTAACACC
 TGTTTCAGCGATAATCTTGTGATGGTTCTCCACCCCTAACGATGACAATCTAAC
 GTCCACATCAATCTTGTATCAATTTCGGAGCAGTTGGAGCCAATTCTGGACGAAC

TTCTGGAATGGTTGCTTCAATGACATCAAGGATTCAAAACGCGTTCTGGCTTGAGC
 AAGAGCCTCCGTCAAGATTCTGCAGTAATCCCTGAATCTGATATCATTGAAGGGC
 TGTAATCCCATCACGAGTACCTGCAACCTTGAAGTCCATATCTCCAAAGTGTCTTCAA
 ACCTTGGATATCTGTCAATACTGTGTAGTTATTCCATCTGAGATAAGTCCCAGCAAT
 ACCAGCTACTGGCGCTTGATTGGCACACCACAGCCATAAGGGCAAGAGTTCCCGACA
 GATAGAAGCTTGAGATGAAGAACCGTTGATTCCAAAATTCTGCTACTAGACGGATAGC
 GTATGGGAATTCTTCCAAGCTTGGCAAGACTTGAGCAAGAGCACGCTCACCAAGGGACC
 GTGACCGATTTCACGACGACCTGGCGCACCGTAACGACCTGTTCCCTACAGAATATTG
 AGGGAAGTTAGTGGTGCATAAAGCGTTCTGTACTCTGGATCCAAACCATAATGAT
 TTGAGTTCTCCATCGGAGCCAAGGTCAAGACTGAAAGAGCTTGAGTTGCCACGAGT
 AAAGAGACCTGAACCATGTACAGAGGAAGGAAGTCAACAAACGCATCCAAAGGACGGAT
 TTCATCGACCTTACGACCATCAGGACGCACCTTGTCTGTAAATTAAACGTCGCACCTC
 TCGGTGTTCCATTGTTCCAAGATTTCAGCCACATCACGCATAATACTGGTCAAATTCTC
 GTGGTCCGCATATTTCTCGAAACGGCAGTCACTTGGTCTTCACTGCTTGAGTTGC
 AGCTTCACGGGCCAATTCTTACTTGAACGGCAGTCACTTGGTCTTCACTGCTTGAGTTGC
 TGCAATGATTTCAGCTTCAATTACGACATCCACGTGAAGCAATTCCACTCTGCTTTTC
 TTTACCGACAGCAGCAACGATTCTTCTTGAAGGCAATCAATTCTTGACAGCTCGTG
 CCCTTAAGAACGCTTCAACATGATTCTGTACAATTCTTGACCGAGACTCTAC
 CATGTTGATAGCGTGGTCCAGCTACTGTCAATTCAAGAACGAGATTGCTCTGCTTG
 TTCTGACTTGGTTGATGATGATTGGCCATCTACATATCCACTGTACCCAGCAAT
 TGGTCCGTCAAATGGAATATCTGAAATAGACAGTGCACAGATGAACCAAACATAGCAGC
 CATTGGTGCAGATGCATTTCATCATAAGAACGACTGTATTGATGACTTGGACTTCATT
 ACGAACACCTCCGCAAACATAGGACGAATCGGACGGTCAATCAAACGCGCTGTCAAGGT
 CGCATCTGTTGAAGGACGTCCTCACGTTCATAAAGCCACCAGGAAACTCCAGCCGC
 ATACATTTCTCGTAGTTGACTTGGAGTGGAAAGAAATCCCAGTTGCCATTCTT
 AGACATAACGGCAGCAGTCAAGACAGTTGACTCACCGTAACGTACGACAACAGATCCATT
 TGCTTGCTTAGCAACCTGACCAAGTCTACAATTGATCACGACCCGAAAGTCGTTG
 AACACTGTTGCCATTAAATCCCCTTGGATTGATGAAATTATACGCCTTG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	1	1128	R	376 aa

[SEQ ID NO:] 3864496-6 ORF translation from 1-1128,
 direction R
 VKDQVTAVYEEKYADHEEFDRIIMRDVAEILEQMEHAEVRRLITEDKVRPDGRKVDEIRPL
 DAVVDFLPRVHGSGLFTRGQTQALSVLTLAPMGETQIIDGLDPEYKKRFMHHYNFPQYSV
 GETGRYGAPGRREIGHGALGERALAQVLPSELFPYAIRLVAEVLESNGSSSQASICAGT
 LALMAGGVPIKAPVAGIAMGLISDGNNYTVLTDIQGLEDFHGDMDFKVAGTRDGITALQM

DIKIQGITAEILTEALAQAKKARFEILDVIEATIPEVRPELAPTAPKIDTIKIDVDKIKI
VIGKGGETIDKIIAETGVKIDIDEEXNVFYLLLVDQNAINPCPKLLLWVFREPCKWMKFT
VLNRSYREFGAFVTL*

Blastp and/or MPSearch Result:

Description:

polynucleotide phosphorylase (pnp) homolog - *Haemophilus influenzae* (strain Rd KW20)

Assembly ID: 3864514

Assembly Length: 1678bp

[SEQ ID NO:] 3864514 Strep Assembly -- Assembly
id#3864514
CTCATGTTGATTTTAAACCAAGAAAAACTGCTAATAGTAAGTAAGGATAAAAAGAAA
TAGTATGCTATATAAGAGAAAAAAATCTATAAAGAAACTAGCATTGTTGCAATACTT
ATACCATAAAATTCTCTTAAAAATCAACCTCCTTATCTCCAAAGAGAAAGCTAAACCA
TTACTAAATGCAATCAGAAAATCAATAAAATAAGTCGCCGTCAAATCCCCGTACTA
AGAGCTGCTAATTGAAACTAAAATGGTAAAGTGCTTAATTGATTCAGACGAATACGA
CACTCCAACCTATTAAAATAGTTATTCAATAAAAGAATAATATATATGTGAAC
GGAAAGCAATATACTCCAGTCGTATCTGAAGTAAAATCAAGATCCATTCTAATACA
TTGGATGGATTGAATATTGGCGACAGCGCAATAATATACTGTACTAGATAAACACAG
GATAGCAGTAATATAAAATAACCAATACTGATAAAAATCTTTGTAAAATTGAACA
AATTGTTCAATTACATAGTCCTCTGAATGTAGAAAAATGTACCAATAACCAAC
AACTAACAAATAAAATAAGCAAGATGCCACTAACTAAGGAAAGACTGATATCTTCT
GATATCCAAAGCTAATGTTGTCACAGGTTCTAAGTAAGATAGCCCTAAATAGCCAAA
AAATACCACCAACCACATATAGGCAACTGGGATGAAAATAGCTCCTATTTTTCTTCA
CTAGCAAAGCACTAGCTAGTCCAAAATAGAGAACACAGGCCAAACTCCATACCA
GAGTCGTACAAGACTATAGAGCAACTGATTAGAATCAAATAATTCTTAAGGCACCAC
TATAATCTCCAATATAAATTCTCTGATAAGGAGTCACTAAAAGATTAATTCTAATAATA
ATAAAATAGGGAGAAAAAGACTAGAAAAGAAGAAATAATTGCACTACTACCTACAATAG
CCAGATACTCTTTAGAAATTGGCACAATTGCTGTTAGAAAATGACTCTCAGCAT
CCTCTATTATCTGACTAGAATAGGGCAGTGTACAGATAAGTCAGCTACTAGGCTAATCG
GTGAAAATCCTCGAATAGAAGAAGCGGCAAAAAAAATCGATAAACCTCAATTATAAA
TACCATTGAAAGCAAGGAAATTCCCTAAACTCATGCAAAGAAGGGTCAAAAATAAGACA
TATAAAATCGAGGTGATTGAACGACTCCGTACAAGATTACAAATGAAAATTCCATCCTT
ACTCCTCCTTATAATAAAATAGGGTGTAGCATTCTTTCATGCTACACCCACAATCA

ACCATCTTAAGGCTTACTCTGACAAGTAAGTTAATAAGAACTGGACTCCAAGAACCTG
AAGTATGAATTCTTACATGATTCCAAATTGTGGGCCATAGCTAATCTAGTACCAAGAAC
CAATATAATTGTCACCACCTCCATTATAGTACATGACAATCCTAGAGGCCAGACCCAATG
AATATACCGGGTAATATCTGACCCACTATAGGCGCTACGAATAGAGGTACTAACCTT
TACCGCCACCAGTGTCACTGTTATTCAATTCCAGCAGAGGCCTTCTTCTCAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	551	937	R	129 aa

[SEQ ID NO: 3864514-6] 3864514-6 ORF translation from 551-937,
direction R
VTPYQEIYIGDYSGALKELFDSNQLLYSLVTTLWYGVWGAVFSIFGLASALLVKKKIGAI
FIPVAYMMVGGIFWAILGLSYLEPVTTLALGYQKDISLSLVSGLAFLFVSVCLVVYGT
FLHSEDXV*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3864518
Assembly Length: 2908bp

[SEQ ID NO: 3864518] Strep Assembly -- Assembly
id#3864518
CTGGTGAAGTTGACTGAGACCGAAGCGATAGGCATCCATGATAATCAAGACAGTCGCACT
GGGAACGTTGACCCCAACCTCAATAACCGTCGCGAAACCAGAATATCCGTCTTCTCTC
CTTGAAATCCTGCATGATCTGGCTTTCTGCACTCTCATCCTACCATGTAAAAGAGC
CACCTCTGTCTCGCCTGCAAAATGAGTCGTAACTCCTCTGATAAGGCAATGGCATT
CAAATCTAGAGCTCTGATTCTCAATCAAAGGAGAGATGACATAGACTGGAACCTT
TTGAATTCCCCCTCTAACCAAGTCAAGACCTGAGGTAGTTGCTCATGTTGATCCAGCG
CGTCACAATAGGCTTCCGACCTGCTGGCATCTGGTCGATAATGGAAACATCCATATCTCC
AAAGGCTGTGATGGCAAGCGTCCGTGGAATGGGAGTCGCCGTATCATGAGGACATCTGG

ATTGTCGCCCTTTCCCGTAAAATACGCCCTTGCCTACACCAAAACGGTGCTGCTCATC
GATAATAATCAAACCAAGACGAGCATACTCCACCCCATCTGTATCAGAGCGTGAGTCC
TATAATCAAATCAGCCTCACCTGGCAATGGTCTCCAAGACTTCTCTTTCTGCAGC
TTTCAAGGAACCTGTCAAGAGAGCCAGTTCAAATTGGAAAAAGGTTCTGTAAACTCTC
AAAGTGTGCTCTGCCAGGGATTCTGTGAGGACTTCTGAGGACTTCTGAGGCTGATAACCTGCTGT
CACTGCCGAAACATGCCAAGCCAGCGACTACCGTTCCGCTCCCCACATCTCCTTG
TAGGAGACGATTCATGTGGTGGCTGGACTTCATATCAGTAAAATTCTGCAAACACTCTT
TTCCTGAGCTTGGTCAGGGAAAAGGAAGACTTGCTTAACTGCTGTCACTTTCTG
AGACCAATCCAGAACCAAGACCCTCCCTGAACCTCTATTTCAAGACTTGAGCGTCTGCAG
CTGCATTTGAAATAAAAGAGTCCCTCAAACCTGATACGGCGAAGAGCCTGCTGTATT
TGCCAAATCCTTGAAATGCATAGCTGGACTGCCTGACAACGGGACATGAGTTGTA
TTTGTCTAGTAAAGACTGGGCAGATTTCTTATCAAGAGGTCCAGTCCCTGATCAA
AGCCGCTTGTGATGACCTTGACCAGACTGGCCTGACTGATTCCCTGAGCCAGACGATAGAC
AGGCTGGAGGTTCATCTTACCTGAGCCAGAACCTTCATCCCAGTCAGACTAGCCTTAGC
GCCGCTCCATTTCAAAAGACAGCAAGGGTTGCTCCAACTCTATTTCAGCCAGATA
GGCCTGGTAAAGAAATTCAACCGAAAAACGACCTCTCCCTGCTTGAGACTAAAACGCAG
GCGATTGCCTTGAACCAATAACTGGACACTAGCAGGAGTCACTACCTGACCAGAAAG
AACTGCCTTCTCACCGTCTTCTAGTTCCAGCACCTGCTGGTTTGAAGTCTTCATAACG
GAAAGGAAAGTAGAGCAAGAGATCTTGCAAGTTCAATTCTAGTTGGCGTATTTTC
TGCTGACTTGGTCCCACACCAGGCAAGACATGCAAGGGTTGATGTAGATTGATGCTCCA
CTCCTTCTTTCTAATAATATTCTCGGAATACGGTCGCTGAGGAGGCAAACCAACCTC
ATAGTTAATGGTACCGGGTAGGTCGCTACCTGAGTTGCAGTGATTCCCTTATCCCCATT
GGAGCCAATCAAGGTACCTGGTCTAGCGGATAAAAGCTTAGGCAATCGAATAGTGTAT
TTGGTCCATCGAAACCTGCCGACAATTGGCAAGCTTGGCCATCTACCAAGACAGAGAA
ATTTCGCATGTCCTGTCCATCCATCTGCATACCGATTGGCACGGTCGCGATGACTTG
CTCGCTATCCGCTTGATAAGTTGCTCCATAGCCATGCAAGCTCCAGCTGGAACTGTCTT
GACATGAAACCAAGAGCAGACTCCAAGGTCAAGGCCGTATCAAATCATAAGGCAAATTCA
AGACCGCTCCACTTGGATTGAGGCCATACATGGCATCTCCATACGAACCGCATTGAAAA
TAGTCTCTACATGCCAAAAGTCGTTGAGAATTGCTAGCATGAACCAGCTCTGGAACCTT
CCTTCATACTAGCTAAAATAGTATTAAACCGTTCTAACTGGCATTAAAATAGTCATCTG
ATTCCCTCATCAGCAGTAGCAAAGTGGTAAAGATTCTCAACACGAACACCGTGTGTT
GGGAGCAAATCTTGAGCCTGCTCAACCTCACTGGCCTCTCTAAAACCAATCCGCTCCATC
CTGAAATCAATCTTGAGGTGGACTGTCAATCCAGTTAGGTCACCTCCTTATCTAAGAGT
GCTTGGAACTCCACTCCAGCCACTGTCAAGGTGAAGTCATATTCTTAGCTAGAAG
CAACAGCTTGTCTGAGTTCAATGGCTCATCAAACCTCTAAAATGAGGATTGGCTGCTG
AGTCCAGCTGTCTGAGTTCAATGGCTCATCGATATTGGAAACGCAAAAGCCATCAACA
TCATCTTGAAATTGCCTTGGCAACGGCAACAGCTCCATGGCCATAAGCATTGGCCTTGACC
ACAGCCCACTTGAGCGTTCTGAGGGATATGAGCCCCCATTGCTGAATATTGTGCGA
ATAGCTCCAGATGAATCAGAACCTGGTTGGCTATGTTGGACTAACTTCATGATT
CCCTCCAAAATGACACTGGCTGTCACAAACTGATCGGTGGCTGAATAAACAGCCAAA
TCTTTCTGAAAATGGTGGCCTGACT

ORF Predictions:

ORF #	Start	End	Direction	Length
8	1985	2371	R	129 aa

[SEQ ID NO:] 3864518-8 ORF translation from 1985-2371, direction R

VRLSLRKICSQQHGVRVEGIFTHFATADEESDDYFNAQLERFNTILASMKEVPELVHASN
SATTFWHVETIFNAVRMHDAMYGLNPSGAVLNLPYDLIPALTLESALVSCQDSSWSLHG
LWSNLSSG*

Blastp and/or MPSearch Result:

Description:

ALANINE RACEMASE (EC 5.1.1.1). - BACILLUS STEAROTHERMOPHILUS.

Assembly ID: 3864522

Assembly Length: 1549bp

[SEQ ID NO:] 3864522 Strep Assembly -- Assembly
id#3864522
CCAGTTAAGGCTGGTTGTCGTCCTCTGGTAAAGAGAACCTCCTTGTAGAGCCTGCAT
TAATAAACTTACGAATGGTTCACGAGCAGCTTCATAAGGAAGCTGTCGCTCGTCCGCT
AAGGTATGGACACCACGGTGAACATTGGCATTGCTGCTCATAGTAACTGTTAATAGCT
TTCAGAACTACTAGTGGTTTGTGTCGTCAGCATTGTCAGATAGACCAGAGGTTCA
TCATTGACAATCTGATCTAAAATTGGAAAATCCTGCGAATCGCTTCTACATCTAACATA
GGCTTCCCCTTAGCGTTTGACAATTCTCTCGATAGTTGCAATCATTCTACGAAAC
TTCCTGACTGGAATCTCCACGATAACAGATCCAAGGAAACCACGAACAACCAAACGCTC
TGCAGTTGCCTTATCCAATCCACGACTCATGAGGTAATACATGTCTCTGGATCAACTTG
TCCGATAGACGCTGCGTGTCCAGTCAGCATTTCTACATCAAAGAATTGGGTT
AGCATCTGAACCGCTGGTCTGAAAGCATGAGAACACGGCTCTGTGCGCATCTGC
TCCCTTAGCACCCTTGATGATGTGGCCGATACCATTGAAAGTCAAAGTTGCTTTCAAG
GATAACCCATGTTGAGGATATTCCGATAGAGTTGCGAGCCATAGTTAGTTACACGAGT
ATCAATCCCTGTACCTGACGACCACTTGAAAGAGCTACAACCTTGAGGTCAGCATGGCT
ACCATTACCAATCAAGTCACTATCAAATCAGCAACGACATTCCTCGTCATGACACC

GATAGCCCAGTCATACTTGCATCGTGCCTAATTCCATACCACGACGGCTAATGTAGGC
 AGTGACGTTTCACCTAGACGGTCGATAGCAGCAAACCTGACTTGCACCCAGAACGTGC
 AATCACTTCCACTGTGATATTGGCAGTTACTTGTCACTTCCTCACCGCGTACTCTAA
 ACGCTCCAGATAACTAATCTTAGAATTTCACCAGCGATAATCATAATATGCTTGTAAA
 CGGCACATTGCTATCGCTATCTGGTAGAAAATTCCCTCAATTGGCTCTGTGATTCTAC
 GTTATCTGGAATATAGAGTACAGCACCCTGTTAAAGTAAGCTGTGGTAAGCCGCCAA
 CTTGTCATCATCATACTAACAGATGACATGAAGAATTCTCGATCAGCTCTGGAATTTC
 TTCTAAAGCTGAGTGAAGTCTGTGAAGACAACACCCCTGTCAGCTAACTCAACTGGAGT
 TTGTTGAAAACAGTTGAGTTCCACTTGCACCAACTCAAGTGTGATCTAAAGCTGT
 GAAATCTGGAACATTGCTGATGGCTCATTCTGTAACTGTTCCATCACCCAAATTCCA
 ACGGGTGGAATTGACACGCTCAATAACTGGTAATTCCAAAGTCTCAATCTGGTCAAAA
 AGCTTTTGACGGAAATCAGCCAACCAAGCTGGTCCAGCGGTGCATT

ORF Predictions:

ORF #	Start	End	Direction	Length
7	310	1458	R	383 aa

[SEQ ID NO: 3864522-7] 3864522-7 ORF translation from 310-1458, direction R
 VSNSTRWNLDGTTGNEPSANVPDFTALDHHLKLVQVGTQTVFEQTPVELAEQGVVF
 FD FHSALEIPELIEFFMSSVKYDDDKLAAYHTAYFNSGA
 VLYIPDNVEITEPIEGIFYQD SDSNVPFNKHIMIIAGKNSKISYLERLESRGEGSDKVTANITVE
 VIARSGAQVKFAAIDRLGENVTAYISRRGMELGNDASIDWAIGVMNEG
 NVVADFDSLIGNGSHADLKVVVALSSGR QVQGIDTRVTNYGCNSIGNILQHGVILE
 KATLTFNGIGHIIKGAKGADAQQESRVLMLSD QARSDANPILLIDENDVTAG
 HAASIGQVDPEDMYYLMRG
 LDKATAERLVVRGFLGSVIV EIPVKEVRDEMIATIEEKLSKR*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864568
 Assembly Length: 1548bp

[SEQ ID NO: 3864568] Strep Assembly -- Assembly
 id#3864568

CTTGGTAGAACTTGCTAATCAAGCTGGCAAGCCTGTAGTCTGGACTGCTCAGGTGCAGC
 ACTTTCAGGCTGTTCTGAATCACCCATAAACCAACAGTCATCAAACCAAATAATGAAG
 AATTGTCTCAGCCTTCTTGGAAAGAGAAGTTCTGAGGATTGGATGAATTAAAAGAAGTA
 CTTCAAGAACCTTGTTCAGGGATTGAATGGATTATCGTTCACTGGTGCAACCG
 TACTTTGCCAACATGGTGACACTTCTACAAGGTAGATATTCTAGAATTCAAGGTGGT
 AAATCCTGTTGGATCTGGAGACTCTACTGTGGCAGGAATTCTTCAGGACTTCTTCACAA
 AGAATCGGATGCAGAATTACTCATCAAGGCAAATGTCCTGGTATGCTCAATGCTCAAGA
 AAAAATGACTGGTCATGTCAACATGGCCAACATCAAGTTCTATATGATCAATTAAATAGT
 AAAAGAGGTATAAAATGGTTAACAGAACAAAACGTGCACGCTAGAAAAACTTCTG
 ATGAAAATGGTATCATCTCAGCTCTGCATTGACCAACGTGGTCTTGAAACGCCCTCA
 TGGCTCAACACCAACAGAACCAACTGTGGCTCAAATGGAAGAACTGAAAGTCTTGG
 TAGCAGATGAATTGACTAAATACGCTTCATCAATGCTTCTGACCCCTGAGTATGGACTTC
 CAGCAACTAAAGCTCTGATGAAAAAGCTGGCTTCTCCTGCTTATGAAAAAACAGGTT
 ATGACACAACAAGTACAAAACGCTTGCCAGACTGCTGGATGTTGGTCTGCAAACAGTA
 TTAAAGAAGAGGGTGCAGATGCAGTTAACATTCTGCTTACTATGATGTAGATAGTTCA
 ACGAACTCAACCAAGAAAAACAGCTTATATCGAGCGTATCGGTTCTGAGTGTGGCTG
 AAGATATCCATTCTCCTGAAATCCTGCTTACGATGAAAATCGAATTGAGACGCCAG
 GTTCTGTAGAATATGCGAAAGTAAAACACACAAAGTTATCGGTGCTATGAAAGTCTTT
 CAGACCCACGCTTAAACATTGATGTCTGAAAGTTGAAGTTCCTGTTAACATAAATATG
 TTGAAGGCTCGCTGAAGGTGAAGTGTTTACACACGTGAAGAAGCAGCAGCAGCAG
 AAGCGCAAGATGAAGCAACGAACCTGCCATACATTACTTGAGTGCTGGTGTATCAGCTA
 AACTCTCCAAGATACTCTGTATTGCTCATGAATCAGGTGCAAACATTAAACGGAGTTC
 TTTGTGGCCGTGCTACATGGCAGGATCAGTTGAAGCTTACATCAAAGATGGTGAAGCAG
 CAGCTCGGAATGGCTCGCACAACTGGATTGAAAACATTGATGAGCTCAATAAGTTC
 TTCAAACACAGCGACTTCATGGAAAGAACGTGTGAAGAAAGTCCCTAGTTAGGAA
 CATGAATCTAAAAAAATTCAAAAAAGTTGTATGTAAAGGTTACAAA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	296	493	F	66 aa

[SEQ ID NO: 3864568-6] ORF translation from 296-493,
 direction F
 VVNPGSGDSTVAGISSGLLHKESDAELLIKANVLGMLNAQEKMTHVNMANQVLYDQL
 IVKEV*

Blastp and/or MPSearch Result:

Description:

TAGATOSE-6-PHOSPHATE KINASE (EC 2.7.1.-)
(PHOSPHOTAGATOKINASE). - LACTOCOCCUS L ACTIS (SUBSP.
LACTIS) (STREPTOCOCCUS LACTIS).

Assembly ID: 3864590

Assembly Length: 1360bp

[SEQ ID NO:] 3864590 Strep Assembly -- Assembly
id#3864590
CTTCCTCCAGCAAAATCCACTGCTGAGAAGCTAAAGGGAGCGTGAGATAGCCCTCTTCT
CTACTGGTTGGTCTGAAATCCGAGCCTCAGGAAACCAGTCTGTAGTTCTTTCCCTCA
TGTCTAGCCCTCCACTTTGGATGCACCATGAAACCAAACCTCTCAAGACGTTCCAGAT
TCTCAGTCATATGGAGATAGCCCATAACCGCTTCAAATCCCGTGGACATACGATAAGTCA
CGACATCTGCATTTAGCCTTGTGTGGCTATTGGTATTGGCGGCCACGTTGTAGATT
CTTCTCTTTTCCGTTAGGACCTGCTCCTCCAACATGAGAGCAATCAGGCGAGCCTGAG
CCTTGGCTGACACATACTGGTTGCTTCTGATGGAGTTATTGGGTTGGTCATACCTT
TGAGGATGAGGTGACGGCGAATATACATAGAATAACACCGCATCCCCCTCAAAGGCTAGCG
CAATCCCGTTAATGAGATTGACATCAATCACGTGTCACCTCACTCCATCCTGGTATCA
AGGAGCTTAATTCTTGAGTAACCAATTGGTCACGGATTGGCTGCTGTCTCAAAGTCT
CGATTGGCACGCGCCTCTGGCGTTTGAAATCAAGTCTCAATCTCTGCATCCAAAAC
TCCTCAACAAAGACAATTCCAAAATTCTAACATATCTGCAAGAGCTGCTTGACACTT
GCATCATAGTCCCTGAGTTGATCCATTGGCATTCAAAGACAACTGTGATACCGTTG
GCAGCATTAAAATCTCATCCATAGCTGCTACAAACTTATCTTAAAGTTGTAACTCT
TGGGCATCCACGTTCTGTAAATGGTTGTCGTAAGTATTCTTCAGATACTTGAGATTG
GTCTCGGCATCGCGAAGTGCCTTCCGTGAAGTTGATAGGCTTACGGTAGTGTCTGGGTC
GCAAAGAAGAACGAAAGTACTTGCCTCATCAAGAGTTAAGGGCATCGTGTACCGTAATG
AAGTTACCCAAAGGACTTAGACATTGGCATTGTCGATATTGACAAAGCCATTGTGCATC
CCAGTTAGTTAGCAAAAGCCTTGCCTGTTAGCTTCAGATTGGGCAATTTCATTGGTGT
GGTGTGGAAACTCTAGGTCAAGCTCCACCACCGTGGATATCAATGGTATCACCTAAAATCT
CTGTCGACATGACTGAACACTCAATATGCCAACCCGGACGTCCAGGTCCCCAAGGACTAT
CCCAAGAAATCTCACCTGGTTGGAAGATTCCATAAAGCAAAGTCTACAGGATTTCCT
TACGAGCCGTTCTTCATCGGTACGACCTGAAGCACCTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	125	511	R	129 aa

[SEQ ID NO:] 3864590-6 ORF translation from 125-511,
 direction R
 VIDVNLINGIALAFEGDAVYSMYIRRHLILKGMTKPNKLHQEATKYVSAKAQARLIALML
 EEQVLTEKEEEIYKRGRNTNSHTKAKNADVVTYRMSTGFEAVMGYLMHTENLERLESLVS
 WCIQKVEG*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864596
 Assembly Length: 2130bp

[SEQ ID NO:] 3864596 Strep Assembly -- Assembly
 id#3864596
 TTGACAAACGGTACTTATGTAGTAGGGACAGCACTATCGGAGCAGGAGCGGTCAATTACCAAT
 TCTATGATTGAGGAAAGTAGTGGTGCAGACGGTGTGACAGTCGGTCCTTATGCTAACAT
 TCGTCCAAATTCAAGTCTGGTGCCAAGTTCATATTGGTAACTTGTTGAGGTGAAAGG
 ATCTTCAATCGGTGAGAATACCAAGGCTGGTCATTGACTTATATCGGAAGCTGTGAAGT
 GGGAAAGCAACGTTAATTTCGGTGCTGGAACTATTACAGTCAACTATGACGGAAAAACAA
 ATACAAGACAGTCATTGGAGACAATGTCTTGTTGTTCAAATTCAACCATTATTGCACC
 AGTAGAACTTGGTGACAATTCCCTCGTTGGTCTGGTTCAACTATTACTAAAGACGTGCC
 AGCAGATGCTATTGCTATTGGTCGGTGTGTCAGATCAATAAGACGAATATGCAACACG
 TCTTCCTCATCCTAAGAACCGTAGGAGCCTATCATGGAGTTGAAGAAAAACGCT
 TAGCCGAAAGAAATCTATCAAGGACCAATATTAAACTGGTCCAAGATCAGGTTGAATT
 ACCAGAACGGCAAGGGAACTGCCAACGGGATTGATTTCCACAATGGGCTGTCTGT
 TTTAGCAGTAACGGATGAACAAAAACTTATCTTGGTCAAGCAGTACCGCAAAGCTATCGA
 GGCTGTCTCTTACGAAATTCCAGCCGGAAAATTGGAAGTAGGGAGAAAACACAGCCCCTGT
 GGCAGCTGCCCTCGTGAATTAGAGGAAGAAACAGCCTATACAGGGAAATTAGAACTT
 GTACGATTTTATTCAAGCTATTGGCTTTGTAATGAGAAGTTAAACTATATTAGCAAG
 CGATTGACAAAAGTGGAAAATCCCGTCCGCAGGATGAGGATGAAACCTTGGAAAGTCCT

TGAAGTGAGCTTAGAAGAAGCGAAAGAATTAATCCAATCAGGTATTTGTGATGCCAA
 GACAATTATGGCTGTTCACTATTGGAGTTGCAGAAAAATAGAGGAGGTCAAGTATGGGT
 AAATCTTATTAAACGGATGAAATGATTGAAAGAGCTAATAGAGGCGAAAAATTCAGGT
 CCTCCTTGCTAGATGATAATGAGGAAACTAAGATTACCAACCTCTTCCCGTTT
 GGTTATGCCAATCCTAAGGATCATGGTTTAGCCAGGAAACCTTGAAGATTCAAGTCGAA
 CCATCTATTCAAAAGCCGTCGTATTGAAAATACCAAGAGAAATGTCTCAATTCTAAG
 TTGAATAAAATCTTATTGCGGTATCTTCTCTTGATTGCTTGTAGCAATGAAA
 CTTTGTAAATAGAAAAGGATTGAAATGAAAATAGGAATTATTGCTGCTATGCCAGAAGA
 ACTGGCTTATCTGGTCCAGCATTAGATAATGCCAGGAGCAAGTTGTTGGGAATAC
 CTATCATAACAGGAACCATTGCTCTCATGAAGTCGTTCTTGAGAAAGTGGATTGGTAA
 GGTCATGTCTGCTATGAGTGTGGCGATTGGCTGATCATTTCCAGGTGGATGCCCTTAT
 TAATACGGGTTTCAGCTGGGCAGTAGCAGAAGGTATCGCTTGGGATGTCGTGATTGC
 TGACAAATTAGCCTATCATGACGTGGATGTCACAGCTTGGCTATGCTTATGGACAAAT
 GGCGCAACAACCGCTTATTCGAATCAGACAAACCTTGGCTCAAATCCAAGAGAGT
 TTATCTCAATTGGACCAAAACTGGCATCTTGGTTGATTGCTACAGGAGATAGTTGTT
 GCAGGAAATGACAAGATAGAAGCGATTAAGTCCCATTCCAGAAGTTTAGCCGTGGAG
 ATGGAGGGGGCAGCTATTGCTCAAGCAGCGATGCCCTCAATCTCCAGTCTAGTCATC
 CGAGCTATGAGTGACAATGCCAACCATGAAGCAAACATCTTTTGATGAGTTATTATC
 GAAGCTGGACGTCGCTCTGCCCAAGTCTGTTGGCTTTGAAGGCTTAGATTAAGCG
 GAAATTGACAGTTCTAGATCAAGCTT

ORF Predictions:

ORF #	Start	End	Direction	Length
11	1915	2097	F	61 aa

[SEQ ID NO:] 3864596-11 ORF translation from 1915-
 2097, direction F
 VEMEGAIAQAAHALNLPVLVIRAMSDNANHEANIFFDEFIIIEAGR RSAQVLLAFLKALD
 *

Blastp and/or MPSearch Result:

Description:

PFS PROTEIN (P46). - ESCHERICHIA COLI.

Assembly ID: 3864624
Assembly Length: 2128bp

[SEQ ID NO: 3864624] 3864624 Strep Assembly -- Assembly
id#3864624
ATCGAATTGAGTTGAGGCTGGATAACTATATCCGTATGTTAAAGATCCTGTCTTT
ACAAAATCTCTGATTAACACAGTTATTTGGTTATTGGATCTGTACAGTTGTTCTA
TTCTCACTCTTGTAGCATCTCAGACCTATCATCAAAATGTCATTGCCAGATCCTCTAC
CGTTTCGTCTTCTTCCTGTTGAACGGGTAGTGTGCCGTGACAGTTGTTGGAAA
TGGATTATGACCCACTATCAGGGATTCTAAACTTGTCTTAAGTCAAGCCACATC
AGCCAAAACATTCTGGTGGGAGATAAAAACGGCATTGATGGCGATTATGATTATT
CTCTTGACCACTTCAGTTGGTCAGCCATCATCCTTATATCGCTGCCATGGGAATATT
GACAATTCACTGGTTGAAGCGCGCGTGTGATGGTCAACTGAGTTCAAGTTTTGG
GAAGATTAAATGGCCAAGCCTCTTCCAACAACTCTTATATTGCAATCATCACAACAAT
TAACTCATTCCAGTGTTCGCCCTGATTCACTGGTACATCTTACGAAAAAGCCTTCAATT
AACAAAGTACCTTGATGTACTACCTTACGAAAAAGCCTTCAATTGACAGAAACGGCTA
TGCCAACACAATTGGTGTCTTCTGGCAGTCATGATTGCTATCGTAAGCTTGTCAATT
TAAAGTACTTGGAAACGACGTAGAATACTAAAGAAAGGAGACAGCTATGCAATCTACAGA
AAAAAAACCATTAACAGCCTTACTGTTATTCAACAATCATTGCTCTTGTGACTGT
GCTGTTCATCTTCCATTCTACTGGATTGACAGGGCATTCAAATCACAAACCTGATAC
AATTGTTATTCCCTCAGTGGTCCCTAAAATGCCAACCATGGAAAACCTTCAACAAC
CATGGTGCAGAACCCCTGCCTTGCATGGATGTGGAACCTCAGTATTATCTCATTGGTAAC
CATGTTCTTAGTTGTGCAACCTCATCTAGCAGGTTATGTATTGGCTAAAAACGTT
CTATGGTCAACGCATTCTATTGCTATCTTATCGCTGCTATGGCGCTTCAAAACAAAGT
TGTCTGTACCATTGGTACGTATCGTCAACTTCATGGGAATCCACGATACTCTCTGGC
AGTTATCTGCCTTGATTGGATGCCATTGGTGTCTCCTCATGAAACAGTCAGTGA
AAATATCCCTACAGAGTTGCTGAATCAGCTAAAATCGACGGTTGTGGTGAGATTGTC
CTTCTGGAGTGTAGCCTCCGATTGTGAAACCAGGGTTGCAGCCCTGCAATCTTAC
CTTCATCAATACTTGAATGACTACTTCATGCAGTTGGTAATGTTGACTTCACGTAACAA
TTTGACCACACTCACTGGGTTGCGACCAGCAGGCTGAAATGGCAACCAACTATGGTT
GATTATGGCAGGAGCTGCCCTGCTGTTCCAATCGTCACAGTCTCCTAGTCTTCCA
AAAATCCTCACACAGGGTATTACTATGGGAGCGGTCAAAGGATAAACTCTCGA
CGAATGCAAACACTACGTCAGCTCACCTGCCATACTTAAGTATTGCCGTGGTAGCTC
CTAGTTGTTCTCAATTTCATTGAGGTATAGGAAAATCAATCTATCAAGATAACAGAAG
TATATTATAGATTAGAGAATATAGAAGTTATAAGTGTCTACAAAATGGAGGGTATGC
AGTTACTTTATGAAGTTGTCAGACACTTAAACTTAAGAATGGTTTAGTTAACTAT
CAGAAAACGAAGGAAAGAGTATGATTGACGATTGAAAACATCACCTTACAAAG
GGATTCACTCCAATTAGACAAGGCTATCGACTATCTCTACCAACATCGTAAAGATT
TCGAATTAGGAAAGTATGAGATTGATGGAGATAAAAGTCTTCTAGTTGTTCA
TGCTCAATCAAGTTGAGAATAATCAATTGAACACCATAAGAACTATGCAGATTG
TGCTGATAGAAGGGCATGAATATTGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	446	751	F	102 aa

[SEQ ID NO:] 3864624-6 ORF translation from 446-751,
direction F
VLMVQLSFKFFGKIKWPSLLPTTLYIAIITTINSFQCFALIQLLTSGGPNYSTSTLMYYL
YEKAFQLTEYGYANTIGVFLAVMIAIVSFVQFKVLGNDVEY*

Blastp and/or MPSearch Result:

Description:

MULTIPLE SUGAR-BINDING TRANSPORT SYSTEM PERMEASE PROTEIN
MSMF. - STREPTOCOCCUS MUTANS.

Assembly ID: 3864630

Assembly Length: 1773bp

[SEQ ID NO:] 3864630 Strep Assembly -- Assembly
id#3864630
ATCGAATTATATATAAAATCTTACACATTAGAAAAGGAGGTTCCCATGTACTTCAA
CATCCTCTGCCTTGATTGAATTCTCATCTTGGCCGTACTGGAGCAGGGTGATTCTTATG
GTTATGAGATTAGCAAACCATTAAGCTAATCGCTAATATCAAAGAATCCACACTCTATC
CCATTCTCAAAAATTGGAAGGCAATAGCTTCTGACAACCTATTCTAGAGAGTTCCAAG
GTCGCATGCGCAAATACTACTCCTTGACAAACGGTGGTATAGAGCAGCTTGTACCTAA
AAGATGAATGGGCACTCTACAGACACCATAATGGCATCATAGAAGGGAGTATCGGCC
ATGACAAGAACTGAATACTGACTCAGCTAGAACTCTATCTCAAGAAACTACCTGAAGCT
GACCGTATCGAAGCCATGGACTATTCAGAGAGCTTTGACGATGCTGGAGTCGAAGGA
GAAGAAGAACTCATCGCTAGTTGGAACTCCCAAAGAACGGCCACGAAGTTCTATCCA
ATCTTCTCGATAAAAATCAATGAAGCACCGCTAAAAATAACCGACAAATTTAC
ATATCGCCTTGTAGCCCTCCTGCAGCACCTATCGGCATTCCTCTGGAAATGCCATCC
TCGTGACCCCTGTTGCAATCCTGTAGCCGCTTGTACTGTCATTCTGGCTTCTTGCAG
TTTCCATACTGGGTATCATCGGCGGATTCCATTAGTTGAAAGTTCACTATCCTCG
CCCAAGCCAAATCAGCCTTATCTGATTTGGTTCTGGTTACTGGCTATCGGTGCTT

CTTCGCTAGTTTACTTGGCATTCTATGTAGCTCGCTTCTCGGTCTACTCATTGTTCTCTGGTACAATTGTTCTTAAAAAGGAAAGAGAGGTAATCAGCATGCGTAAATGGACAAAGGATTTCTCATCTTGGTGTGGTGAECTACCGTTATCGGCTTTATCCTGCTTTTGTAGGTATCCAATCTGACGGGATTAAGAGTCTACTTCCATGTCCAAAGAACCTGTCTATGATAGCCGTACGGAAAAGCTAACCTTGGCAAGGAAGTCGAAAACCTAGAAATTACTCTCCACCAACACACGCTCACCACATCACAGACTCTTCGATGATCAAATCCACATTCTTACCATCCA TCTCTTCTGCTCACCACATGATTTTATCACCAATCAGAACGATAGAACACTCTGAGTCTCACTGATAAGAAAAGTCTGAAACTCCGTTCTCTCTGGATTGGTGGGATTCTTCATATCGCAAGTAGCTACTCTAGTCGTTGAAGAAGTTATTCTCCGACTACCAAAAGGGAGAACTCTAAAAGGGATCAACATCTCAGCCAATCGCGAACACCACATCAAATGCTAGCCTT GAAAATGCGACCCTCAATAACAAACAGCTATATCCTCGAATTGAAGGAAGTCGTATCAAACACAGTAAACTCACAAACGCCAATATCGTTAATATCTTGATACAGTTCTACAGATACTCAGCTAGAGTCAACAGATAATCACTTCCACGCTGAAAATATCCAAGTCCATGGTAAGGTT GAACTGACTGCCAAAGATTATCTCAGAACATCATCCTAGACCAGAAAGAAAGCCAACGAATTAACTGGGACATCTCAAGTAACTACGGTTCTATCTTCAATTACACAAGAGAAAAGCCTGAA TCAAGAGGTACGGAATTAAGCAACCCTACAAA

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
8	663	953	F	97 aa

[SEQ ID NO:] 3864630-8 ORF translation from 663-953, direction F
 VTLFAILVAALTIVLAFFAVSILGIIGGFLFLVESFTILAQAKSAFILIFGSGLLAIGAS
 SLVLLGISYVARFFGLLIVRLVQFVLKKGKRGNQHA*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864654
 Assembly Length: 2307bp

[SEQ ID NO: 3864654] Strep Assembly -- Assembly
id#3864654

CCACCTGGATTTCTAACGTTCGCAAGAACATTAGAAGAACAGTTAGCGAAAAATAGAG
CCTTGGAGAGACGTTACTGAGTCGACTCGAATTCAAAAGTAGAACGCGAGAACAGG
AAAAAGAACGTTGTTAGAGGAATTGACCTTCTGCAGGAATATAGATGTAGGTCAAG
CGAGAGTTCTTCTAGCGGCTACTTGAGTTGGAATTGGTACTACCTCTGTCAATATAT
ATGCTGGTATGGATGATGATTTAACGTTACAATGCACCAATTAAACATGGTATGAAA
CGGCTCGCTATGCCTTGAGCGAGGTATGGTCTGGCAAAATTAGGTGGTGTGAAA
CTCTCAATGGTGGACTTATCATTAAAGGAAAATTAAATCCAACGATTGAAGAACACT
TGGGTGAATTACAATGCCACTCATCCTCTATCCTCTGTTAAGACTTGCTCTGATT
TCCGTAAAACATTAAGAAAAAACATAGAAAGTAAGTATATGGCACTAACAAACACTCAG
AAAGAAGAGTTTCAGACTTCTGATCAGGTTCTCGTCCTTATGCAATCTGTC
CAGATGGGGGATTGCTAGAAAAAGAGGGGCTCGAATTGTTATCTTGCTTGAACAA
GAAGGAGAAATTCAAGTTGCAGCTCTGGTTAGCTTGCCTTATGGCTGGTGGTCTGCA
TATGGAACTCAATTGGGGCCGATTACCCAACAAGATGCTCTCCAGTTTATGC
AGAGTTAAAAGAATATGCCAAGCAAAATGGTGTATTAGAGTTGCTGTAAAACCTTATGA
AACTTATCAAACTTTGATAGCCAAGGTAATCCAATAGATGCTGAGAAAAAAAGTATTAT
TCAAGGTTGACTGATTAGTTATCAATTGATGGCTAACAACAGGTACCCAGGTGG
AGAACCAAGATTGGTTACTATAAAGATTAACTGAATTAACTGAAAAGAGTTGCTTAA
AAGTTTAGCAAAAAGGGTAAACCCCTGGTGGAAAAAGGCTGAAACCTTGGCATTGGTT
GAAAAAGTTAAAACGTGAAGAACTATCGATTAAAGAATATAACAAAAGAACCTCTGA
ACGTAGAGAATATAGTGTAAAAGTTAGAATATTAGAGCATTGATGACTTTGG
AGAACAAAGCGGAGTTCTCATAGCAAGCTGAAATTTCGGAGTATATGAGCAAATTGCA
AGGTGAACAAAGTAAACTAGAACAGAAACTGGACAAGTTGCGACTTGATTTGAGTAAAAA
TCCTCATTCTGAGAAAAACAAACTGAGAGAACATTCTAGTCATTGAAACGTT
TGAAGTTGCAAAAGCAGAACGCGAGACTGATTGAAAACGATATGGAGAACAGATATT
GTTTAGCTGGAGTTATTGTTATATGCCTCAGGAAACGACTTATCTCTTAGTGGT
TCCTACACTGAGTTAATAAGTGTATGCCCTGCACTGCTCAAAATATGTTATGTTG
GAAAGCATAAAACGTGAATACCTAAATACAATTCTCAGGCAAGGGATTGGAT
GGAAGTGTGGTTGCCTTAAACAGAATTAAATGGCTATATTGTACGCAAAGCG
GGTACTTCCGTTACCATCCATCGCCTTAAAATACAAAGCTATCCAGTTACTCAAAAAA
ATAGTAGGACGTTAAGATGAAAAGTCAGTATTAGATTCTTTAGCTTCTTTAGTAA
AATCGAATTTTATTGCTAGAAAGGTGGAGAGAACATGCGCTGGCTTTCTGTTGATAG
GGGCTTCTTTTTGTGTGGCGTTGTTGGCGTCTGGTTGGATAGTTGTGCTCT
TATGTGTGCTTGCTTCGGACTTCTCTGGTATTGAAACGGGGATTTCAGGAGCGCTAA
AGCAAGCAGAACGGTCAGTAAAATTGGTCAACAAAGTATTGACCAATGGGAGAACAG
GGCAACTGCCTAAGTTAACGCCAGACAGATAGTCACCAGCATTCTGAAGGAAGGTGGCCAC
AGGCCTCTGCTCGTATTACCTGGATCCGAGATGGATTACGCTTCAAGAGGCTTATT
TAGAAGCAATCCAGAACTGGAATCAAACCTGGTGCTTTAACCTTGAACCTCGTGA
CTAGTAAGGGGGATTACGGCTACGGAGATAACGACGGAAGCACTCCTGTGGCAGGAGA
AGCGGAAAGTCAAACTAATCTCTTAAC

ORF Predictions:

ORF #	Start	End	Direction	Length
9	1878	2306	F	143 aa

[SEQ ID NO:] 3864654-9 ORF translation from 1878-2306, direction F

VWRLFWRLLWIVVLLCVLAFGLLWYLNQDFQGALKQAEERSVKIGQQSIDQWEKTGQLPKL
SQTDQHSEGRWPQASARIYLDPMQDSRFQEAYLEAIQNWNQGTGAFNFELVTESSKADI
TATEITTEALLWQEKRKVKLIS*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3864658
Assembly Length: 1236bp

[SEQ ID NO:] 3864658 Strep Assembly -- Assembly
id#3864658

TTCCCATATTCCTGTNCTTCACCAGAATTGAGATAATGATTGTATTCTCATTTAA
TGATTGTTCAAATTGTGAAAGATAGCTTCTTGGACGTAACCTCTCCAATTGTTATT
TAAAGAGCTCGCTTGTAAACCTTCTGACTTGATAACGAAATAATGACATCTCCAGC
ATTTACCATATCTCCTCTGACTTATGTAAAGTAACACCTCCCTGAACCAATTGCTGA
TAGGAACCTGTACCTGTTATAACTGAATTCCATTGCTTTACAATATAGTTTG
TATATAAGCTGCCAACCAATGCACCGCTTAAGATAATAGCAGTTGAAATAATGAGAAT
AAACGCAAAAGCTGGTGGTCTCTTATCAAAGAAAATACGAGAATAACGTAATTCTGATT
ATTATATAATTTCATAGGCTTACAATTGGCTCTAAATATCTACTACCATTTTCAGGA
GAAGAATTAAACATAACTGTATAGACAATCCCATCCGTTGAATATCATTTCATAGACA
TATAGATCCAATTAGAATACGCATACTGTAGATACTCTGGACTGTCTTCAAAACGAACA
TATAAACAAATGGAACAGAGATAGAACCTGTACATCATAATGTTACTGTACTGTTGA
GCATTATGAGCTTGAATATAAAACTCAAAATCAGTCGTTATTAAATCCATCATCATGAATA
GTAGTACCAACTTTTACAATTAAATGGACCAAAAATTGCTTAAACAACGTGCAA
TGTGATGAAATTATAATTTCCTAATCAACATCTTCTACTTNGTATCATGTAAC
TTACAGATAACTGACTTTAGTACCAAGTTTTATTATCTTTACCTCTAACTAGCCATA

AGTAACCTCCTCTGTATCTAACACAGCCTGTGACTGAATTGTTGATTCACTTGAACGCT
 CTGCAAACCAACCATTCTAGCATACTTCCATTTCGCCATTAGTTCTTCATGAGTC
 ATATTCCACAATAGTCCCATTTCTAAAAGCATATCTTATCACATCGAAGAATTG
 CAGCCTGTGGCTACTACAATTGTTCTTATCCATTATTTATTAAAGATTAAATC
 AATAATCTGTTCACTAAATGAATCTAAGTTAGAGGTTGCCATCAAATATACAAATC
 AGCTTTACTCAGTAGTGCTCTGCAATAGCCAATCG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	892	1029	R	46 aa

[SEQ ID NO:] 3864658-7 ORF translation from 892-1029,
 direction R
 VEYGTHEELMAKNGKYARMVGLQSVQVNQQIQSQAVLDTEEVTYG*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864664
 Assembly Length: 2124bp

[SEQ ID NO:] 3864664 Strep Assembly -- Assembly
 id#3864664
 CCTCGTTATGCAGATGAACGTTATTCTGTCAAAGAGTCACAAGAATTGTTGATCGT
 AATCTTTTATTACCATTCTGACAAGGAAACCACCTGTATCAAGCCTTATCAGCAGGAT
 TTGGATTTGCCACATGGTCTGGCCTGGATGTTGCCTTGGATTATTATCCGAAAAT
 CCAGCTGAGCGGAAAAACNGGTTCTGAGCCTGATTATTCACTCTTGTGCGCAA
 ACTATTCCAGAAAAGCATGGTCTCATGAAATGGGAAGTCGCATTTACTGGGTTG
 ACTCCAAAATCTCTCCGTTATCGCATCTGGAAAAAGCTGAGAAAGAAATGACTAAGTAT
 GATTGGCTGATTGTGATGGCATTACAGAATTATGCTCAGGTCTGGCTACATGAGAAAC
 AAGTACCCAAATCACATCTTGAAGACAATCTTCTTGCCTTGAAGGAACAGAGATG
 CCTATTCCAATCGGCTATGATGTCTATCTCAGAACTGCTTTGGGATTATATGACGCCT

CCACCAGCAGACAAGCAGGTACCGCATCAGGATGCTGTCATCGCTGATATGGATAAGTCT
 TATACAGAATACAAGGGAGAATATGGTGGCTAAGAAAAAAATCTTATTTTATGTGGTC
 TTTTCTCTGGAGGTGGTGCAGAGAAGATTCTATCAACCATTGTTCAAATCTGGATCC
 AGAAAAGTATGATATTGATATTCTGAAATGGAGCACTTGACAAGGGATATGAATCTG
 TTCCAAAGCATGTACGCATTTAAAATCCCTCAAGATTATGCCAAACCAGATGGTTAC
 GAGCTTTTGTGGAGAATGAGAATTATTCAAGACTGACTCGTCGTTGCTTGTAA
 AAGATGATTATGATGTTGAAGTTCTTACCATATTGAATCCACCACTGTTGTTCTCTA
 AAAGAAGAGAAGTCAAGAAGATATCTTGGATTGAGAAGTATTGAAGAACTCTTAAGG
 ATAGCTCTAAAAGAGAATCACATAGAACGCCAGTTGGATGCTGCGAATACAATTGTAGGG
 TTTCAAAAAGACCAGCAATTCTATCAAGGAAGTTATCCAGATTATGCTCTAAATTAC
 AGACAATCTACAATGGATATGATTTCAGACTATTCTAGAAAATCTCAAGAGAAGATCG
 ATATCGAGATTGCTCCTCAAAGTATCTGACTATCGGACGGATTGAGGAAAATAAGGGTT
 CTGACCGTGTAGTGGAGTGTACGATTATTACACCAAGAGGGAAAAACTATCATCTCT
 ATTTTATCGGGCTGGTATGGAAGAGGAAGTGAAGAACTGAAAGAGTCAAAGAGTATGAGA
 TTGAGGACTATGTACATTCCTGGTTATCAAAAAAATCCTTATCAGTATTATCTCAGA
 CGAAAGTTCTTGTCTATGCTAAACAAGAAGGCTTCTGGAGTGTATGGAGGCCT
 TGAGTCTGGACTCCCTTTATCTCTACGGACGTTGGAGGGCTGAGGAATTATCCAAG
 AAGGACGATTGGACAAATCATTGAGAGCAATCAAGAGGCAGCTCAGGCGATTACTAATT
 ACATGACTTCTGCCTCAAACCTTAATGTCGATGAGGCTAGCCAATTCAACAATTAA
 CAATTACAAAACAAATCGAACAAAGTAGAAAAACTATTAGAGGAGTAGCATGGAAACTGCA
 TTAATTAGTGTGATTGTGCCAGTCTATAATGTCGCTGAGTACCTAGAAAATCGATAGCT
 TCCATTCAAGCAGACCTATCAAATCTGAAATTATTCTTGTGATGATGGTCAACA
 GATGAAAGTGGTCGTTGTGATTCAATCGCTGAACAAGATGACAGGGTGTCACTGCTT
 CATAAAAAGAACGAAGGATTGTCGAAGCACGAAATGATGGATGAAGCAGGCTCACGGG
 GATTATCTGATTTTATTGACTCCAAATGATTATCCATCCAAAGAAATGATCCAGACC
 TTATATAACCAATTAAATTCCAAGAAGAATGCCGATGTTCCAAGCTGTGGTGTTCATGAA
 TGTCTCTGCTAATGATAAAACCC

ORF Predictions:

ORF #	Start	End	Direction	Length
7	675	1727	F	351 aa

[SEQ ID NO: 3864664-7] 3864664-7 ORF translation from 675-1727, direction F
 VVQRRFYQPLFQIWIQKSMILIFLEMEHFDKGYESVPKHVRILKSLQDYRQTRWLRAFLW
 RMRIYFPLTRRLLVKDDYDVEVSFTIMNPPLFSKRREVKKISWIHGSIEELLKDSSKR
 ESHRSQQLDAANTIVGISKTSNSIKEVYPDYASKLQTIYNGYDFQTIILEKSQEKIDIEIA
 PQSICITIGRIEENKGSDRVVEVIRLLHQEGKNYHLYFIGAGDMEEELKKRVKEYEIEDYV
 HFLGYQKNPYQYLSQTKVLLSMSKQEGFPGVYVEALSLGLPFISTDVGGAEELSQEGRFG

QIIESNQEAAQAITNYMTSASNFnVDEASQFIQQFTITKQIEQVEKLLEE*

Blastp and/or MPSearch Result:

Description:
amsK protein - *Erwinia amylovora*

Assembly ID: 3864700

Assembly Length: 1660bp

[SEQ ID NO:] 3864700 Strep Assembly -- Assembly
id#3864700
ATCGAATTAAATCCATAAACAGATTGGTGATTGATAGACGACATTGGACAGTTGCGA
TCTGGCAAGACAGAATGTTGGTCAAACGGCTAACATGGTCTTACGAATAGCCTGAAAG
ACTTCTGGATTCCCTGCTGAATATAGGTCCACAATTGGCTTTTGCCAGATGCTCC
GCTGTTTCAGATCGGTTGAGCAGGGTACTGGAAATCACCGTCGTGATTCAATATGATTC
AGCAGATATTCTCGCATTGGGATGACTCACTTGGGACAAATCAAGCTGGTCTACCAAG
AGTCGATTGACCTTGAGTTGCTGGTCAATGCACTTAATCATCACTTGCTCATGACAGAC
TGGCCTCACGCCAATCAAGTAACGATAGAAATCGACAGGCAGATAGTACATGGTCTTG
ACCTGCTGAAGGGCGTAAAGACAAAGAGATTATCGACATAAAAGATGTTCAAGGAGT
TAGAACTGGCTAGCACGCAACAAATCTGTCCGATAATCAGCGAGTGCATCATGATATAC
TGGCCTTGGAGAAATTCCGACCTGGTCCAGCCAAAATCTGCCAACAGGCAAGACT
GACTCGTAACTCATACTCTTACGAGACTGACCTCCTTCTAGACAAAATTGGTC
ACAAAGACATCCATCTTGACCCCTGCTCTCAAGTTCCTGCAAGGTTCAAGAATTTC
AAGTAGGCACGAGGATCCACCACTGCACTGTCAACTACTTAAAGATGCCAGAAG
CCTCTGCCAAGCCGATTGACCACACCGCCATGGCCTTATTTCTGATAGATGGCTC
TAACGATATTAGGATACTTGCTAGCTAACACTCAGCGATTCCTGAGTCTGGTCTGAG
ACCCGTATTGATAATCAAACCTCCACTGCTCACCACCAACTCAGCGACTCCACAC
AGTAATGAAGATAGGCTGCTGCATTATAGCTAGAAATGGCGATAGACAATAACTCATAA
TCTGCTCCTTAGGGACTGATTTCTTACCTTCGAAAATCTCTCAAACCGCGT
CAACGTCGCCTGCCGTAGATGTTACTGACTTCGTCAAGTCTATCTGCAACCTCAAA
CAGTGTGAGCAGCCCGAGCTAGTTCTAGTTGATCTTGATTTCTGAGTAT
TACTCTCTTGTCACTCCTCTATTTACATAAGTCCAGCCTTGAAGAACTTTA
CTAGAAGACAAGGGCTCTGCTCTATTGCCATCTGGCATCAAAAGAGGGTCA
TCCCTCTTACGAATTCAATGCTACTAGGGTATCAAATACTGGTTGATGACTGCCA
AAATATAGGTATCTGCTTCAAGAGGTATCTGGTCAAATTCAACATCCAATGGGAAT
TTTCTGCTCTCGAAACCCAAAATTCAAGATTGTATTTGCCACGGAGGTCTAATTAC
TCAGACTTTGACCTGCCAAGACTGAGGAATTTCATCTCCACGATAGACACATTTTAT

CCAACTGAAAGACATCAACACTATTATGGAAAAGAATGGTCTGTGCTAGAGACTGCCCA
 TTTCATACTCTGGCGAGATAACCGAGTCAGCTCCCATCTT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	480	740	R	87 aa

[SEQ ID NO:] 3864700-6 ORF translation from 480-740,
 direction R
 VDPRAYLKILETLQELESKGQEMDVFTNFVYEKEGQSRKKSMSYESVLPVRQIFGWDQV
 GNFSKGQYIMMHSLIYRTDLLRASQF*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3864706
 Assembly Length: 1306bp

[SEQ ID NO:] 3864706 Strep Assembly -- Assembly
 id#3864706
 CTGATCGAATTAAAAGAAGCCCACCCCTAATCTGCCTACTTCTTACCTCCAACACTGGT
 CGTGTCCAACCTTATCGAGACATTGACCTGGTGGCTCAAAAAAGGTCAAGATTCACAGA
 CCAGGAAGTTGTCCAATTTATCTAGACCTCTCATTCTCAAAATTGAATATAGAGTAA
 AGCTTCAGTTGTCTTATTCTAGGTTACTGAGTTTTATCTTTCAACAACAAAAGAGG
 ACCCGCCGATCCTCTTTCTACTATAAAATCCTGATTATCAACTATATCTTTAAT
 CGAAATCTCAAAACAGCACTTCAAACATCTTCTAGTTAAGTAAATCAGTATTTGC
 TTAGCTGCCCTGCTCCATTGATACCAACCAACTAGACTGTTAATGAGATAAATTAGATAT
 TTCCCTTGAATTGCAAGGTTCTCCCCACCAAGAGATAGATTGAAAAGACATTGGTAGCC
 GCCCAGAATATCCACTGTTCACGGTAAACAGCTGTCATGAGGATTGCCCTACCCATTG
 GTTGCATCTGTGATTGAATCACGATAGGGACGATTGGCACCAATAGACTGATAATGAA
 GCCAAAGGCAACCACCAAAGCACACTAATGGAAAGATACTTGTCCAGCCCTGCCGTC
 CAGTTACGCGCGACAAACTCCTGCTTTCTTAAACTGTGCCTGATAAATCCAAA

CTAGAGTCCAATTGGCTGCATGACTGTGAAGTAAAGTGTGTCAGCACCTCACCATAAAA
 GCCTTTCTGTAGGGCCAATAAGGTAATAACAGAGTTAACAGCCAAAAGATAATTAC
 TTGCTCGACCTCCGATACAAGATTACACAGATAATCCCTGTCAAGCTACAAATCATCCC
 AATCCAGTCAACAATACGATGTTGTAACCAACTCCAGCCAGAGAGGAAAACCTCCTAA
 AACCAGCAAATAATCCACTGGGCAAACACTACGATGGGCAAAGAGGTCATCCCAGATAGC
 CTTCATAGTTCCTGAAAATCTAAATCAGCCATAGCCGCAACCATACGACGGTAACCACC
 TGACATTTCACCTAGGGTTGTTTGATATTTCAATTTCTTTGCAAATAAGTATGCAT
 CATTCTCCTTTGTTTAAAGAGCCGTCTGGATAGACTTCGGACGCAACGCTCTA
 TTAGATAATGAACTGCCTATACACAAGATTCTAACCTAGTCGACATGAGCTGAAACCT
 CTTATTTGTTAAGTAGTTCACNAATATTACACCTATTTATGA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	336	626	R	97 aa

[SEQ ID NO:] 3864706-6 ORF translation from 336-626,
 direction R
 VCFGWPLASFISLLVPIVPYRDSITDATNGVGQILMTAVYREQWIFWAATNVFSIYLWW
 GESLQIQGKYLIYLINSLVGWYQWSKAQKQNTDLLN*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864710
 Assembly Length: 1676bp

[SEQ ID NO:] 3864710 Strep Assembly -- Assembly
 id#3864710
 AACACGCTTGGCATGGCAGATAAAGCGAGATTTTTGTTTCTGGACTGGCGTCT
 TCTTAATTGTCCTAAATTCCATGATTAAATTGACTAAAAATAATATAAAGTGCTAGT
 TTTACGAATAAAGAAGTATGAAAGTAAATTAGATTATCTCGGTGTTATTTACTGAG
 AATGAATTAAACAGAAGAACGTCAGTTGGCGGAGAAACTCCAGCAATGAGAAAGGAG

AAGGGGAAACTTTCTGTCAACGTTGTAATAGTACTATTCTAGAAGAATGGTATTCGCC
 ATCGGTGCTTACTATTGTCGAGAGTGCTGCTGATGAAGCGAGTCAGAAGTGATCAA
 TTAACTATTTCCGCAGGAGATTTCCGAAGCAAGATGTTCTCAAATGGCGCAGCAA
 TTAACCTCTTCAAGAGAAGGTGTCAGAGGGACTGCTCAAGCAGTAGACAAGCAA
 CCAACCTTAGTCATCGGTAACAGGAGCTGGAAAGACAGAAATGATTATCAAGTAGTG
 GCTAAAGTGATCAATGCGGGTGGTGCAGTGTGTTGGCTAGTCCTCGCATAGATGTTGT
 TTGGAGCTGTACAAGCGCTGCAACAGGATTTCCTGCGGGATAGCTTGCTACATGGA
 GAATCGGAACCTTATTTGAAACACCACTAGTTGTCACAAACCCATCAGTTATTGAAAG
 TTTTATCAAGCTTTGATTGCTGATAGTGGATGAAGTAGATGCTTTCTTATGTTGAT
 AATCCCACGCTTACACGCTGCAAGAAATAGTGTAAAGGAGAATGGATTGAGAATCTT
 TTAACAGCGACTTCGACCAATGAGTTAGATAAAAAGGTCCGTTAGGAGAACTAAAAAGA
 CTGAGTTACCGAGACGGTTCCATGAAATCCGTTGATTATTCCAAAACCAATTGGTT
 ATCGGATTTAATCGTACTTAGACAAGAATCGTTGTCACCAAAGTTAAAGTCCTATAT
 TGAGAAGCAGAGAAAGACAGCTTACCGTTACTCATTGCTTCAGAAATTAGAAAGG
 GGAGCAGTTAGAAGAAATCTACAGGAGCAATTCCAAATGAGAAAATTGGCTTGTATC
 TTCTGTAACAGAGGATCGATTAGAGCAAGTACAAGCTTCGAGATGGAGAACTGACAAT
 ACTTATCAGTACGACAATCTGGAGCGTGGAGTTACCTCCCTGTGATGGGATTTGCT
 AGTAGAGGCCAATCATCGTTGTTACCAAGTCTAGTTGATTCAAGATTGGTGGACGAGT
 TGGACGAAGCATGGATAGACCGACAGGAGATTGCTTCTTCATGATGGGTTAAATGC
 TTCAATCAAGAAGGCATTAAGGAAATTCAAGATGATGAATAAGGAGGCTGGTCTATGAAAG
 TGCTTGTATGTGGCAGACTATGAAGACTGTTAACTTTAGTAGTCTCTTACTTCTG
 AGGAATGATGACTCTGTCTTGTTCAAGTGTGATTCTACTTTGAAAGAATTGGGAA
 GAGAACTGTCCAAATTGTATGAAAACAGAGTTGTCAACAAAGTGTCAAGATTGTCAACTT
 TGGTGTAAAGAAGGAGTTGAAGTCAGTCATAGAGCATTAACTTACAATCAAGA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	442	972	F	177 aa
7	1247	1438	F	64 aa

[SEQ ID NO:] 3864710-6 ORF translation from 442-972,
 direction F
 VSEGLLQAVDKQNPTLVHAVTGAGKTEMIYQVVAKVINAGGA
 VCLASPRIDVCLELYKRL
 QQDFSCGIALLHGESEPYFRTPLVVATTHQLLKFYQAFD
 LIVDEVDAFPYVDNPTLYHA
 VKNSVKENGLRIFLTATSTNELDKKVR
 LGEKLRLSLP
 RRF
 PWKS
 VDYS
 KTNL
 VIGF*

Blastp and/or MPSearch Result:

Description:

COMF OPERON PROTEIN 1. - BACILLUS SUBTILIS.

[SEQ ID NO:] 3864710-7 ORF translation from 1247-1438,
direction FVDVFVVEANHRLFTKSSLIQIGGRVGRSMDRPTGDLLFFHDGLNASIKKAIKEIQMMNKE
AGL*

Blastp and/or MPSearch Result:

Description:

COMF OPERON PROTEIN 1. - BACILLUS SUBTILIS.

Assembly ID: 3864724

Assembly Length: 2159bp

[SEQ ID NO:] 3864724 Strep Assembly -- Assembly
id#3864724
CTGCTCTACCACATCGGATAACAGCATAAGTTCAACTTATCAAAGCTAAAGTGGTT
CAATTCTCCACCCTGGAGTTGAGCAGGGGGCTTTAGATTAGTAACCTGGTTCCCAG
TTGGCAGAACATCATTAAAGACATGGTCCTTCATTACCAACAAAATAGGGTTTTAGGAGC
TGTGGGACAGTCTTACCAACATAACTCAATCACATAAGACTTCGGTGCACCAACTCC
ATGGTCTTCATGGAAGCCAACGCTTAAGTTATCAACTGAACGTTGCTAAAATACCTGA
ATCTCCGAATAGGACACCGACTGAAGCTCTGGATTACTACGATTCCAGTTGTCCAACG
ATTGGCTGGTTGGTTATTGTAGGAAATGAGCTTGTCTTAACATTGAAACTGGGTCGCT
TGGATTGAAATCTGAAGCAAAGGCAAGTGGCAATTCTGAACCGGTCCATTGGTCAGAAAT
GTTGCACCTTGCTCAGTTGAGCAGATACGCGAACATGAAGTTAGTTGTTAATTGAGT
ACCTTCTAACGCACCATTAACTGTAAAGACACCTTCCTTAGCGTATTGCTCTGGACGAAT
CGCATCCCAGTCAACCTTAGCTGATGAAACGTGACCATTGAATCATATGTCCGAACACT
TTCTGGTAATTGTGGTGCTTCTGCGATTGGAGTTGTCACACTGACTTCAACTGAAAC
GATACCTTCTACAGAGACTTTGCACCGCTTCAAGGTCAATTCTCAACTTACCTAG
TACCTCAAATGTCTGATAGGAGTCTAGTTTTCTTGGAAATAGCTGCGAAGTGAAC
ATGAGTTTAGGAAACCTTGTCTACTCAACTGTTACTGTTGCTGGAAAGACTGGTTC
CTGATGCAAATCTGTCACTACATTACAGGACGGATGGATTGCGCAATCTTCTCAGT
ATTGGCTTGGATAGTGTGAGTTCAACTTGGCCTTGTAGCTCCCTCATATTGCGTTCAAAGT
GACTGCTCCTGGCTTATGCAACTCAAGCATTCTTACGAATTGCGACTTCCCCTTCACC
ACTTGTAGAGAAGGTTACTTTACAGCTGGTAATACAGCTTGCCTCCATCTGATAGTG

AGCTCGAACCGACAATTGACAGTTGGTCTTCTTGAGACTGTCAGCTTTCCACTTG
 CAAGCTCAAGTGAGCAATTGGCGCTTCTTCAAGGAATTGAATTGCATAGGTTGAAGA
 GGGCCACCACCTTAAAGATGCTCGCACGCATGCCGTTGCTGCGCTTGCT
 TGAAGAACTGTAACAGCTGCATTTAGCACTTGCTGTGACTTCTGGCAACTTAGCTCCA
 TAAGCAAGAGTGCGGTATTGCATTGGTTTGACTAGTAAGACCTGTGACAGCTTCACCA
 CCAACCGTTACAGTTGGTACTGCAGGTGCCGCAGGATTGCCCTCTTCTACCAAGGGTT
 GCATGAATTGGTTGACCTTCTAAATAACCGGTCGCTTGAATACGAGAACCTGGAATTGCT
 AACTTAGCTTATCTCTCGGCAATCTCCCACCTGTCACCCATCTCAACACTT
 CCATCAGTCAAAACATAGGAAACAGATTGTCTACAGAATTCAAGTCAGTATTGGAGCA
 ATACGTTTACAACAGGTAGCTCTGATTAAGAGCAATCACTTACACGAGCTCTACT
 TCTCGTCCGTCAGCCATACCTTCACCGTTACAATACCAGGCTGTCACATCTACTGAA
 GACCAGGTTACAGGACGTTCTGCACGGCTACCATCACTGTATACAAACGGAACAGTGGTA
 GGCATTTCAAGGTGCCCTCTCCAATAATGGTCTGTACTTTGGCACTTCTGTCCCCAAAACA
 GTCTTCTCTTGTCCCTCTTCTTACCAAGACAGTGACTTGGTTGATTCAAGAGA
 TCAGAGTGGGCAGTAAGGGTGAATTCCCTGTTGTTGACTTGACATTGACAATGGCAACA
 CCTTTACCATTAATGCTTACGAATCCAAGAACCATCTGCTTGCCTTATAGCGTTCA
 CGACTGGCTGTTCTCCGTTATCTACACCGACCAGTTGACCTTGGCCATGCAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	133	1197	R	355 aa

[SEQ ID NO: 3864724-6] ORF translation from 133-1197, direction R
 VEKADSLKEDQTVKLSVRAHYQDGTQAVLPADKVTFSTS
 GEGEVAIRKGMLELHKPGAVT
 LNAEYEGAKGQVELTIQANTEKKIAQSIRPVNVTDLHQEP
 SLPATVTVEYDKGFPKTHK
 VTWQAIPEKLD
 SYQTFEV
 LGKVEGIDLE
 ARAKVSVE
 GIVSVEEV
 SVTTP
 PIAEAPQLPES
 VRTYDSNGH
 VSSAKVA
 WDAIRPE
 QYAKEGV
 FTVNGR
 LEGTQLTT
 KLV
 RVSAQTE
 QGANI
 SDQWTGSEL
 PLAFAS
 DSNPS
 DPV
 SNV
 NDKL
 ISYNN
 QPANR
 WTN
 WNR
 SNPE
 ASVG
 VLF
 GDS
 GILSKRS
 VDN
 LSV
 GF
 HED
 HGV
 GAP
 KSY
 VIE
 YYVG
 KTV
 PTAP
 KNPS
 FVG
 NEG
 PCL*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3864734
Assembly Length: 2199bp

[SEQ ID NO:] 3864734 Strep Assembly -- Assembly
id#3864734
CTTATCGTACTAAGGATGGCAGTGTCACTGTTCCGTCTGATGAAAATGCTAAACGCC
TGCAACGTACATGTGACCGTCCTTGATGCCAACAAAGTCCGAACAGACATGTTGTAGA
AGCTTGTAAAGCAGTGTCCGTGCGAATGAAGAACATCGTACCAACCATACGGAATAGGTGG
AACTTATATCTCGCCCTCTTGATGGTGTGGAGATATTATCGGGTAAAACCGGC
AGAAGAGTACATTTCACCATCTTGCTATGCCAGTTGGAAATTACTTAAAGGTGGTT
GGTCCCACCAACTTCTTGATTCAGGATGAGTACGACCGTGCAGCACCAAATGGTACAGG
TGCAGCTAAGGTTGGGAAACTATGCTGCAAGTCTCTTACCAAGGAAAATGCCAAGTC
ACGCCATTCTCAGATGTTATCTATCTGGACCCATCAACTCATAACAAAGATTGAAGAAGT
CGGATCAGCTAATTCAGCTTGAATTACAGCTGATAATGAATTGTAACACCAATTGAGTCC
ATCTATCTTGCCATCTTACCAAGTATTCTTGCTTATTGGCAGAACATCGCTTGGG
ATTAACTCCTATTGAGGGTGATGTTCCAATTGATAATCTTGACCGTTGTAGAGGCAGG
TGCCTGTGGTACAGCAGCGGTATTCCTCCAATTGGAGGTATTCAACATGGTATGATT
CCATGTATTCTATAGTGAACACAGAAGTAGGTCTGTGACCGTAAATTATATAATGAATT
GACGGGTATTCAGTTGGCGATTGAAGCGCCAGAAGGTTGGATTGTAAAAGTAGATTA
AAATAAACCAAAGGAGATTTTATGAAATAGAAAAAGTGGCTCTAACAGCAGGAGTGG
TCCTGAGCACGTCAGCTATTGAGGGTGATGTTCTGAGACAGTACAGGATGGCTC
CGACAACATTCCTTATGCTATGCAGTAGATCCAGCATCATTGGCTACAGTACCGA
CTCGAACATCGAGGACAGACGTTATTGAAATGTTATTGATGGTTGATGGAAAATGATA
AATACGGCAATGTTGCTCCTCTCAAAAGACTATGATTGAAACAGTACAGGATGGCTC
CAAGCTATCAAGATCCAGCGTCTACTTGAATATTGATCCAAATCTGGTTCTGCCA
TGAAACACCTTGGCATTACGAAAGGAAAAGATAAGGATGTTGAGCTAAACCTGGTTGG
ATAAAATATAAGAAATTGTTAGAAGATGCTGTTCTGAGACCACTGACCTAGAGAAGAGAT
ATGAAAATATGCCAAAGCTCAAGCTGGTCAGACAGATACTCATTATTGATGCCAACAG
CTTCATCTGGTGGTCTCCAGTTGTAAGTAACGTAACGACTACCATTCTCAAAACCATACTCAC
AAGTTGGTATTAAGGGGAACCATATATCTTAAAGGAATGAAATTGCAAAAGATATTG
TTACAACAAAAGAATATAACGAGGTTTAAAAATGGCAAAAGAAAATTGGAATCCA
ATAGCAAATACCAAAAAGAACTAGAAAATCCATTAAATAAGGAATGGTATTGATCTTGA
TAAAATTTCAAAATACTGTCATTTGAATATAAAGGAGTTGATATGGAGTGGATTACA
TTAATAGGAATAGCAATCATTGTTGTTGCTTACAGCTTACAGGTTCTCGTTGAATT
AGTAGTCTTAGCTGGTTGGTTACAGCTTACAGGTTCTCGTTGATGGTTACCTT
GGAGATTTGGGAAAAGAATTAGCAATCAGCGAGTGCTCACGATTGTTATGGTTACCTT
GCCCTTGTGGGCTGTCAGAACCTTGGACTCAAGCAACGATCAATCGATTGATTG
AAAGATTAAAGGTCTGACAGTTGAAACTCTATACAGTTATTCTTATTGAGAGATT
AGCTGGTTCTTCAATTGCTTAGGAGGACACCCCTCAGTTGTCAGACCTTGGTTCA
ACCTATGGGAGAAGCAGCTGCAGAGTCTCAATTAGGTAGAAAGTTAACAGAGGTTGAAGA
TGAGACAATAAGCGCGTGCCTGCGAATGAAAATTGAAATTCTTGTCAAAA

TACGTTTGTAGGTGCTGGGGAGTCCTCTTGATAGGGG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	897	1601	F	235 aa

[SEQ ID NO: 3864734-7] 3864734-7 ORF translation from 897-1601, direction F

VVLSTSAILVACGKTDKEPDAPTTFPYVYAVDPASLGYSIPTRTSRTDVIGNVIDGLMEN
 DKYGNVAPSQKDYLNLSTGWAPSYQDPASYLNIMDPKSGSAMKHLGITKGKDVKVAKPG
 LDKYKKLLEDALSETTDLEKRYEKYAKAQAWSTDTSLLMPTASSGGSPVVSNVLPFSKPY
 SQVGIKGEPYIFKGMLQKDIVTTKEYNEVFKKWQKEKLESNSKYQKELEKSIK*

Blastp and/or MPSearch Result:

Description:

aliB protein - *Streptococcus pneumoniae* (oligopeptide binding protein)

Assembly ID: 3864740

Assembly Length: 1118bp

[SEQ ID NO: 3864740] 3864740 Strep Assembly -- Assembly
 id#3864740

CTCCTATTGGTATTTGCGAAAATTTCTCCATCAATCCAGTCTGGATAAAGACCAATAG
 TCCAAACCCAAAAGTAGGAAGACTGAGCCACCTAACAGAGTAGACTGAAGGCAGACAGATA
 AAGAACCATCACAAATGAGGACAAGAACATGGCTAACATGAGGAAGAACCAAGGAAAGTTAAA
 ACTAGCCAACATCAATCCTTTGAAGAATTCTTCAAGATAGGTACAAACGTGCCGC
 GATAGGGTAACTAGCCAGCATCACGATAGTAAGAAAAATCAGAACATACCTAACAAATGGC
 TTTCAGCAATTGGAAGGGCAGAGCTGTTGACCCAGAAAAGATAGAGATCTGAAAGGGT
 AAGAAACACAATTCTAACTCCATTAAACCCAGCTGAAGACCTAGTTCAGATTTGCTT
 GAAAGATCTAGATAGATTTAAAAACAGGCACCGTCTGCTCTTCAACTCGAACAT
 GGTCTCGTAGAGGCTGATTTAGCCACTCCAATCGTCACGATGGTAAACAAGAGACGAC
 AAAAAGAAGATTGGCTGTCACGATGTCCAAGACCTTCTCACTAAAACGCATGAGAAAGTT

ATCTGTATCAAATGCTGCCTTGATAAGGCTTACTCCTTTGTGCCATGTTGCTCCTCC
 ATCATTCTTCTTGAACTGTTCTTTGTCAAGCTTCATAAGTCCCTACC
 ATGAACAAATCTATTTCTTTCTTTGGACTTTCTATTTTATCTATGGATAT
 ATAATGTATATATAGCGAGGACAACGCACTAGCTAAAATATTACGCCAAGTGTGTTCATC
 AAATCCATTATTCCACGGATTATCATTGCAAGCACTGTCCAAGCTAACATACAA
 TAAAAAAATACAAAGTGCCTTCATTCTGCATTTAAAAGTTATACGACCATTGTTAGGG
 ATTTTATCATGTGCATCCAAAGCTGCAGCAATATTGTAGGCAAAATTACCATACATCA
 GCTACATTCACAGCTATTGTAAAATCCTCCAGAAATCTGGTCAGTAATCCTACTCTT
 GCTGCTGCAGTTGCAGCTGCCCTACTTAAGATCGATCG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	4	264	R	87 aa

[SEQ ID NO: 3864740-6] ORF translation from 4-264,
 direction R
 VMLASYPIAARYDLSWKEILQKGLMLASFNF PWFFLMLAILVLIVMVLYLSAFSLLLGG
 VFLLFGFGLLVFIQTGLMEKIFAKYQ*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3864792
 Assembly Length: 1431bp

[SEQ ID NO: 3864792] Strep Assembly -- Assembly
 id#3864792
 TCCAAATAAGGAAAATAACACTTCTCAAGAAAAACACAACAAGAAGAACGCCAAAATC
 TAGCGTCAAGGAAGAGAAAAAAATCAGAAAACCAGCAACTTCAGGACTCTAATAACA
 CCTGCTACAAGTAAACCTGCCACTGAAAATGAAAACAGGCCAATACTCCAATTTCAGAA
 AATAATACTCAATGAAAATCAAAGAGCAAATAGGAAGGCTAGCCGTAGGCAGTACTTGAG
 TACGGCAAGGCAAAGCTGACGTGGTTGAAGAGATTGCGAAGAGTATAAAAGTAATCAA

TAGCCAGTAAAATAGCTCCTCCAACCTGGAAAGAAGCTATTTTATTGCTGCAATAC
 TTTCTGGCTTGGTACCTCAGCTGGACCAATGACACCTGCCATCTCAAGCTCTTCAT
 GAGACGGGTCGCACGGTTAAATCCAACCTGACAAACGACGCTGAATCATGGATGCACTGGC
 TTTCTGTGTTCGATAACCAAAGACTTAGCTTCTCAAAAAGCGGATCACCACCAGCATC
 TCCATCCGAAAATTCTCCTCATTTCAGAAACCTCACCTGGATCAAAACTCTCATCGTA
 GTCTGCATCTGCCTGAGTCTGATGAAGTTACAATGCCTAACATCGTCATCCGAGAT
 AAAGGAGCCTGGAGACGAACGGATGATTTCATTAATCGGTTAAAGAGCATGTCCTCC
 TCGACCAAGAAGTTTCTGCTCCATTTCATCCAAAATCGTACGGGAGTCTGTTCTGA
 TGAAACCGCAAATGCTACACGGAGATGGAACATTGGCCTTAATCAAACCAGAGATGACATC
 AACAGATGGACGCTGAGTTGCAAGAACATCATGGATACCTGCAGCACGCGCCTCTGCC
 AAGACGGATGATGACATCTCCACTCCTGCTGGCCACCATCATGAGGTAGCCAACCTC
 ATCCACAATCACGACAATGAATGGTAGCGGAATTGCTTGTACTCAGACTGGGAATCGAA
 CTCGTCTACCTTGGCATTAACCGAACTCTGCAACAGCCGAACCTCCACCTGGCAAAGAGTT
 ATAACGGTTGCCATTTCATCCACAACCTTTGCACAGCCCTGCTGGCTTGCCTGGATT
 GGTCAACCAGTGGCAATCTAACAGGTGGGAATATCACTGTAGAACAGATAACTCAACCAT
 CTTTGGGATCGACCACCCATCCTCAGTAAATTAACTTGATCTGGCTCGCCTTCATGAG
 AATGCTANCAATAATGCCGTTAAGTACTGCTACTGACTTCCCTGAACCGTTGAACCTGCAAC
 TAGCAAGTGGGCATTTAAAAGGTCAAAGCTCTGCGGTTCCATTAACAGCCTTCCC
 TAAAGGAATTCCAAGAAATTCTGCTTGCATTGTCATAGTT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	346	1149	R	268 aa

[SEQ ID NO: 3864792-6] 3864792-6 ORF translation from 346-1149, direction R
 VVTNPRKASRAVQKVVDEMANRYELFAKVGVRNAVAGFNAKVDEFDSQSEYKQIPLPFIVV
 IVDELADLMMVASKEVEDAIIRLGQKARAAGIHMILATQRPSVDVIISGLIKANVPSRVA
 FAVSSGTDSTILDENGAEKLLRGDMLFKPINENHPVRLQGSFISDDDVERIVNFIKTQA
 DADYDESFDPGEVSENEGEFSDGDAGGDPLFEEAKSLVIETQKASASMIQRRLSVDNFRA
 TRLMEELEMAGVIGPAEGTKPRKVLQQ*

Blastp and/or MPSearch Result:

Description:

STAGE III SPORULATION PROTEIN E. - BACILLUS SUBTILIS.

Assembly ID: 3864830
 Assembly Length: 1412bp

[SEQ ID NO: 3864830 Strep Assembly -- Assembly
 id#3864830
 AGACAAATCTGATCAATCCGTGGTCGGAAACTCCAAAGTATGTGCTTTATGTTCAAGG
 GATACAGGGCTTGGTAAATCTTCCGTTCGCGTCAACCCCCATTAAAGCCAGAGCTAG
 CAGTCGGGTCAATTGATACAAATTCTATAATTCTCTCTCATCTGCCACTGCAGATAGT
 AGGCCTCTTCCAGCGCCCTCTTTAATAAAGTCAGAATTCTGTCTTCGCGTCAAAA
 GATTTTTTGACGTCTAAATTATTTAGCAAACGGTATTCCCTCCGAGCTGGTATCAG
 ACATTGGGAGAGTTCTTCATTGATGACTCTCACGGTCTACAAGACGAG
 TTCCAACCTCTCTCCAAGCTGACTGAGTTGCAGTCTGACTATTAAATAAAAGGTA
 CACCGAGTACAGATGCAAATAAAAGTAAGATAATCCAGTTAAACGACTTTGAAAACCT
 TTTCAATAAAATAGACTAACATCTTCATCAAACCTCTTCTATCTGCCCTGAT
 GAATGGTTACTACTCTATCGCAGATATCAACCAACTCTCCTTATAGTGGAACTTAAAA
 GAACCAGCTGTTCTGTCTATCGATTGTCTAGCCTATCAAAAACCTCTGTCTATAAT
 ACTCGTCTAACGCCATTGTAATCTCATCCATGAGCCAGCATTGGCCTGACTGAGAAAAT
 ACATAGCAATCACCAAGCGTTGCTTCATCCCTAAGGAATACTTGCGGATGGGAAGACTGA
 TATAGTCAGCCATTCCCAGTAGGCGATTTCATCTCTCAAGTTAGGTCTGACTCCAGA
 TGTTTTTATGAGACGAAGGTAGTCCATCCCCTTAAGTTCCATCCAGCCATTCAACGC
 TCTCATAATAAAACAAAGAAGGAGGAACCTGCGATGTGTCCACTACTAAGGGAAAGCAACT
 TGCTCATAGCTCGGAATAGTGTGCTTCCCAGGCCATTGATAGCAAGAAGGCCATAAA
 TCCTACCCTTTAAAGGTAACATATCCAGCCCTCCTTCTCACTCTTAAGGATTAATAA
 CCTCCAGTATAGTAGTTATGACCTCATAACGAGCGTAGTCCAGCCCTCCGCCAACTTTAAGG
 TAACATGAGTAAGATTAAACATATCCAGCCCTCCTTCTCACTCTTAAGGATTAATAA
 TACTCAGAATAGCTGTAATAACGAGACCATTCCGGAACTCAAGCATACTGATGGCTGTGA
 TAGTTGGTACTATATTCCAAAACCGTATTCCAATCATACTGTAACTTTAGTGGCTGTC
 ACAGCAGATACACTGGACTGAAGAATACCAATAGATTATAAACTAACTAATAAAACAACT
 TTTGCTGATTTTAATGATTTATATCCTCAA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	515	1123	R	203 aa
7	1134	1322	R	63 aa

[SEQ ID NO:] 3864830-6 ORF translation from 515-1123,
direction R
VRKGGLDMLNLTHVTLKTRQVILQDADFTFKKGRYGLLAINGSGKTLFRAMSKLLPLS
SGHIAVPPSLFYYESVEWLDGNLSGMDYRLRIKNIWKSIDLNLRDEIAYWEMADYISLPIR
KYSLGMKQRLVIAAMYFLSQAKCWLDEITNGLDEYYRQKFFDRLAQIDRQEQLVLLSSHY
KEELVDICDRVVTIHQGQIEEV*

Blastp and/or MPSearch Result:

Description:

ATP-BINDING PROTEIN BEXA. - HAEMOPHILUS INFLUENZAE.

[SEQ ID NO:] 3864830-7 ORF translation from 1134-1322,
direction R
VTATKSYKYDWNTVLEYSTNYHDHQYAWIPEWSRYYSYSEYKVGGWNYARYEVINYTG
GY*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3864848

Assembly Length: 1640bp

[SEQ ID NO:] 3864848 Strep Assembly -- Assembly
id#3864848
CTAACAAAGGTATGATACCAAGCACTAGCCAAGGTAGCATTAGCTTCTGTACCTGTGTTG
GCAATTCCCTCTCTTACCTGTCTCATAGTCGGAACCTCTGGGTCTGGATTCACTGGAG
TTTCAGTTTGGAGTACCTGGTTCTGGAGTTGGTTATCTGGTGTGATAAACGGTCAT
ACCTTACCGTTATTCTTATCACTAGAGTCTGACGTAACCTCTGTGATTCAACTGTTG
GAATATCTGGATCTTGTACTTGTCAATCTTACCAAGATATAACCTCGTCCCAGTTCTG
TTGTCCATTACCGTAGGTTACAACCTCCGTGACCTTGTCTCAGTTTGTACGGCTTA
AGGTTACAGGTTGAACAAACATCTTCTTACATTGGTTCTGTAACTTATCAACGTAAT
GAATGATAACGCGTTATAGTCTTCGTCTCAGTAGAGGTTGCTGTTGGGAACCACTGTT
CCTCAACATTCTCACGGTAGTAATAGTCAACTGTTGCACCGTCTCTGGTACGCATTG

AGGAGTTGCTACCAAGGTGTATGTTGTTCCCTTGATAACTCGGTCTTCTTGTCCTC
 AGTGTGTTGTTCCCTCAATAGTTTGATTCTGTGGTATACTCAGAACCTATCGCTAA
 ATCAGCTTTATAACAGACTCTGCCAACCTCTGGCTACCTCTTATAGTAATTGCA
 TGTTACTGTAGCAGTGGTGGCGCTTCGCTTACTCTATAAACTAAGGTCACTGTTCTAC
 CTTCGCTTACAATATTCCCAGTTAACTGCAGAATTGTATCTGCTTCTTAAAGTAT
 AATATTTCCGTCAGTAGTGTACTGCTACTGAGTTTTATCTGTGACATAATAGCTGG
 TACCAATCAGTTGTTTTATTGTAATGTAGGTTCCGTACTTCTTTCTCCAATT
 CAGTATCATTTCATGATAGCAAACGCCCTGTTCATCAACATAGCGAACCTTCACAT
 TTTCTGAGATTAGTTCTGCCAATTCTGAGGTTTTCTTGATTTCTCGGTAA
 TTTCCCTTCTCTCGGGAAATTAGTTGGAATGATTTTCAACAAACGGTTC
 GTGATGGTCCACAGTATCTGGATGACTGAAAAGTCAGCTAGAATTGGGAGATTATAAT
 GAACACGGTGACTTGAGTGTACTCCTACTCTTCATTATTCTCTGAAAATACTCGTA
 CGGTATAAGAAACACATCTTCCTAAATAGAACATCCCCAGTAGAGAAAATAGCCGCC
 TTCCTAGTTGCTATCCAGAGTCCACTTCTCTTAATCTTATCAGATAGTTTTAC
 CAGTCAGTACATTGTCGACAATCCCTTGTCTACCCCTACAAAGTGGGAGAACTTT
 TGAACCTTCAGAACACCAGATCTAGCCCAACCATTATTAAGGGCATTGCTTTGTATTG
 TATTCTCTCTCAAAGGTTGGCGATTAGAATTATATTCACTGGCACTTAGAGTTGCTGCT
 ATATCTGACTCTTGAATACCAACTCCTACTACCATTCTAGCGGCAGTATATGTGAAT
 TAATCTGTTTAACTTCTAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	707	1546	R	280 aa

[SEQ ID NO: 3864848-6] 3864848-6 ORF translation from 707-1546, direction R
 VPMNIILIAKPLRENTNTKANALNNGWARSGSEEFKKFSHVGVVDKGIVRNVLTGKKLS
 DKIRKEVDGDSKLGKGGYFSTGDVLLGKDVSYTVRVFSENNERVGVNTQSHRVHYNLP
 ILADFSVIQDTVEPSRTVVEKIIPKLNIPEEKKGKITEEIKKKKTSSELAELENSVKVR
 YVDEQGRLLSLKNDTGIGEKEDGTYITNKKQLIGTSYYVTDKKLSSMTTDGKYYTFKE
 ADTNSASLTGNIVSEGRTVTLVYRVKRSANHCYSNIELL*

Blastp and/or MPSearch Result:

Description:

MURAMIDASE-RELEASED PROTEIN PRECURSOR (136 KD SURFACE PROTEIN). - STREPTOCOCCUS suis.

Assembly ID: 3864878
 Assembly Length: 861bp

[SEQ ID NO:] 3864878 Strep Assembly -- Assembly
 id#3864878
 CTGGGGAACTCAAATTGTTAATGTTATCATCAAGGGCGGATGTAACAAGGTTATGTNGG
 AAGCCTTCTGCCTCAACTCAAAAAGATTGAACGTGGAAGGTGTCAAAGTGAATCGT
 CCACTCAGCGGTGGTGCATCAACGAATCAGATGTGACCCTGCCGAAGCTTCAAATGC
 CTTTATCGTGGTTCAACGTACGCCCTACACCACAAGCTCGTCAACAAAGCAGAAGCTGA
 CGATGTGAAATCCGCTTCACAGCATTATCTACAAGGTTATCGAAGAGATGGAAGAAGC
 TATGAAAGGGATGCTGATCCAGAATTGAAGAAAAGTTATTGGTGAAGCGGTTATCCG
 TGAAACCTCAAGGTGTCTAAAGTCGAACTATCGTGGATTATGGTTATCAACGGTAA
 GGTTGCCCGTGACTCTAAAGTCGTGTTATCCGTGATGGTGTGTTATCTATGATGGCGA
 ACTCGCAAGCTTGAACACACTACAAAGATGACGTGAAAGAAGTGACAAACGGTCGTGAAGG
 TGGATTGATGATCGACGGCTACAATGATATTAAGATGGATGATGTGATTGAGGGCGTATGT
 CATGGAAGAAATCAAGAGATAAGATTGGCTCCTTCTTAGGTGGTGAGGGACGCAAG
 CAAACCGATGGTTTCATTGCTTATTTGAGCCTAGGGTCTAAAAATCCCCTGTGATGG
 GACTGATAAATCAGTCCATCACTTCACCACGGCAAAGAAGCAGATGACTCAAATTG
 AACATCGTTCAATTAAACTGAAAATCAAGAAGTTAAAATAGCTAGGTCTGCTGGCCT
 AGCTTTGGTCAAAGTAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	95	622	F	176 aa

[SEQ ID NO:] 3864878-6 ORF translation from 95-622,
 direction F
 VEGVKVTIVHSAVGAINESDVTLAEASNAFIVGFNVRPTPQARQQAEADDVEIRLHSIY
 KVIEEMEEAMKGMLDPEFEEKVIGEAVIRETFKVSKVGTIGGFMVINGKVARDSKVRVIR
 DGVVIYDGELASLKHYKDDVKEVTNGREGGLMIDGYNDIKMDDVIEAYVMEEIKR*

Blastp and/or MPSearch Result:

Description:

INITIATION FACTOR IF-2. - ENTEROCOCCUS FAECIUM
(STREPTOCOCCUS FAECIUM) .

Assembly ID: 3864950
Assembly Length: 1469bp

[SEQ ID NO:] 3864950 Strep Assembly -- Assembly
id#3864950
ACTCTTCAGGAATAATTGCATATGTTGAAGACAAATCTCAAACAACTTAGCCTTT
ATTATACTGTAAGAAGATATAGTTCAATTATAGTTCTCTAAGTAGTTAGTCTA
TTTTATATCCTAGTGTAAAGAAAACAGCCCTAGGGACTGTTTCATTAATAATGCATAA
GAACTTGTAGTCGTAGTCACCAATTTCACGGCCGTTCAATTCAATCCAAATTCAACA
AGGAAGGCACAACCTGCCATAACACCACCAAGTTTCATCATCTCGATAGTTGCCCTA
ACAGTTCACCTGTCGCCAAAGGTCTACATAAGAACACGTTGACCTGGCTTAATG
GCATCCGGCGTGCATAGTTCAAGGTATTGACACCGTACTCTTTCATAGTCAGCAGAA
ATAACTTCGCGTGGCAATTACCTGGCTTACGAACAGGGCGAAAACCAATTCCAACTCA
AAGGCAACTGGACAACCCACGATAAATCCACGAGCTTCAGGGGGGTCCCACGGGGGAT
CATGCCGACTTCTGGTCAGTAGCATACGTGAACGATCCACGGGGAACAGGAATTCGT
AGCTATAAGCATTCCATCAGCCATCAAAGGACTAATATCACGGAAGGTAATGCCCTCCT
TTGGATAATTTCATTGTCATGTAATCTTTAAATTCACTTTCTTTCTTCAA
AGTTTTTACTCTATTATAGCATTTTAAGAAAGAAAAAGGAAAAGTTAACCTC
AATAATTATCTAACGTTTGACGATTATAACTAGCCATCGCAATAAGGCCAATTCTG
TTTATTCTTAGCAAACATTATACATAGTTAAAAGCTGCTTCTATTCTCCTTTACA
AGCATTACACAAATTCAAAGTTCCTAGCAAACCTCGTCATAATCATACCCGATAA
TTTCATTAATGTCATTCCACAGTCATGCTTCACATCACATAACCTGATTCTATCAT
CACCTGTTCCCAACCATCTTGAGTTAAAGGACCTACATTACATGAATTGCTTGTGATAA
TTCTGTCTGATAGACTCTTAGCTTCTTAAGAAGCACATCATGTCAGAGAGAAC
TCCAGGTTTAATACCCTTAGATATTCCATTACACATTCTTAGCTGATCGGCTTG
CATAGTCAGCATAGCTTCATTATAACAATATCAAAGACTAGCATCTGATAAGGAAGTTT
CATTGCATTGCTTTCAAAGTAAATGAGCAACACCTGCCGTTCCAGCAGATT
TTTAGCCACTCTAAAGCTTGAGCATCCATACACAGCAGTTATCTGCAACCAAAAG
CTGTGCCAACTCAATTGCTGTAGTCCCCTATTACACGCAACCTCTAGTATTCTTTTC
TTTGAAATCCTCTGCAATTGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
-------	-------	-----	-----------	--------

6

198

500

R

101 aa

[SEQ ID NO: 3864950-6] 3864950-6 ORF translation from 198-500,
direction R
VGCPVAFELGIGFAPVRKPGKLPREVISADYEKEYGVEYLELCTPDAIKPGQRVLIVDDL
LATGGTVKATIEMIEKLGGVMAGCAFLVELDELNGREKNW*

Blastp and/or MPSearch Result:

Description:

ADENINE PHOSPHORIBOSYLTRANSFERASE (EC 2.4.2.7) (APRT). -
ESCHERICHIA COLI.

Assembly ID: 3864954

Assembly Length: 1073bp

[SEQ ID NO: 3864954] 3864954 Strep Assembly -- Assembly
id#3864954
CTAAATAGGGTATAATATGGTAATCATTGTCGTAGGTTGTCTGAAATATTGTCCAG
ACAAGGCTCACAGCAGTTAAATCTTCTGAAAAGTCAGATTAATAGCTGCTTTTGT
GCTTTTTCAAGATTTGAGCATTGTAACAGAGGCTAAAGATTCTGAAAATCGTCA
AGAGGACACGGTATAAGGGTTACAACCATATGGCGATTAGAAAAGCCTGATTGACAA
GGCTTGGAACTTATTACAAAGGAGAACATCTTGGCAGGACATGACGTTCAATACGGGA
AACATCGTACCCGTCGTAGTTCAAGAACAGAAGTTCTGACTTACCAAATTGAA
TTGAAATTCAAACGTACTCATCAAAGCTTCAGACCACGGCTTAAGGAAGTGTGTTG
AAGATGTATTGCCAATTCAAACCTCACAGACACAATGGAGTTGAAATTGTTGGATATG
AAATCAAGGAACCAAAACACGCTAGAACAGAAGCTCGTATCCACGATGCTAGCTACTCAG
CACCAATTGGTAACCTTCCGCTTGATCAATAAGAACAGGCAGAACAGACCCAAAG
AAGTTTCTTGGTATTCCAAATCATGACAGAACATGGAGTTGAAATTGTTGGATATG
GTGAACGTATTATCGTTCTCAGTTGGTCCGCTACCCAGGTGTTACTTTAACGACAAAG
TAGACAAAAATGGTAAGGTGGCTATGGTCAACTGTTATCCCTAACCGTGGAGCTTGGT
TGGAACTTGAAGCGACTCAAAAGATATCACCTACACTCGTATCGACCGTACTCGTAAGA
TTCCATTACAACCTGGTCGTCTGGTTCTCAGGTGATGATGAAATCTTGATA
TTTTGGTGACAGCGAATTGGTTCGCAACACTGTTGAAAAAGATATCCACAAAGAAC
TGGACTCTCGTACAGACGAAGCCTTGAAAGAAATTACGAACGCCTCGCTCAGGTGAGC
CTAAGACGGCTGAAAGCTCACGTAGCTTGGCTCGCTTCCCTGAACC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	414	1070	F	219 aa

[SEQ ID NO: 3864954-6] 3864954-6 ORF translation from 414-1070, direction F

VFEDVLPISNFTDTMELEFVGYEIKEPKYTLLEARIHDASYSAPIFVTFRLINKETGEIK
TQEVEFGDFPIMTEMGTFIINGGERIIVSQLVRSPGVYFNDKVDKNGKVGYGSTVIPNRG
AWLELESDSKDITYTRIDRTRKIPFTTLVRALGFSGDDEIFDIFGDSELVRNTVEKDIHK
NPMDSRTDEALKEIYERLRPGEPKTAESSRSLLVGSLP*

Blastp and/or MPSearch Result:

Description:

DNA-DIRECTED RNA POLYMERASE BETA CHAIN (EC 2.7.7.6)
(TRANSCRIPTASE BETA CHAIN). - BACILLUS SUBTILIS.

Assembly ID: 3864962

Assembly Length: 902bp

[SEQ ID NO: 3864962] 3864962 Strep Assembly -- Assembly
id#3864962

GAATTGAGTGTAAAAGAAATATGAGGATCCCTTAGGGATAGTGGTAAGTAATACCAAAGT
CTCTTAAAGAGGCAAGTGACGAGTCAGAGCAATAAGGCTGAACAACGTGAAAGCCAGC
GTCTTAGGCCTGGCTGATGATTTGGCTTATAGCTCTGAGATAAACCAACCCGTTAGAC
AGGTGGTTATGATTTATCTGAGTGTAAACATACTGTTGGCAATCTGCTGATGCGGTCA
AAGTTGCCTGGGAAGCGAGTTATTGAGTTCGCCACCAATTCAAACGGCGTCTGCACCA
GCAGCGAACCATTGAGGGATGTTGTTAGACCGACTCCTCCGGTTACCATTACGGAAACT
TGTGGGATCGGTGCCTGACTGCAGAGATATGCTGGACTGAGAGTACTACTTGGGAAG
AGTTTGATGATTTCACTACCGGCTTCAAGTGCAGTCGTGATCTCTGTGAGGGTAATACAG
CCTGGAATGTACGGTGTGCTGTAGAGATTGCACATTTCGCAGTTTCAGCATGGAAAGAT
GGAGAAACAACGTAATTGCTCCGGCTAGAATGGCATCTCTAGCAGTTACGGCATCAAGC
ACAGTACCTGCACCGATAACAAACACTCTTATCGCCTGATACAAGTCTACAAGTTCCCTTG
ATGATTGTCCTGCATACTGATTGGTATAGGCGATTCAATAGCTTGATACCGCCCTTG

ATACAAGCAATCGAGGCTTGCAGTCCTCTTCCTTGTATTCAGGGAAATGACAGCGACA
 ATTTTCGATGTTTTAGTTCAATAATCGTATCTGATTGGTCATGTAATTCTCCTAAC
 GAATGATATCTTGTGCATTTGCCAGTAAATTTCAATAACTAGTTGCCAGTGGAGAGAT
 GG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	195	602	R	136 aa

[SEQ ID NO:] 3864962-6 ORF translation from 195-602,
 direction R
 VLDAVTARDAILAGANYVVSPSFHAETAKMCNLYSTPYIPGCITLTEITTALEAGSEIIK
 LFPSSTLSPAYISAVKAPIPQVSVMVTGGVGLNNIPQWFAAGADAVGIGGELNKLASQGN
 FDRISEIAQQYVTLR*

Blastp and/or MPSearch Result:

Description:

2-keto-3-deoxy-6-phosphogluconate aldolase (eda) homolog -
 Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3864970

Assembly Length: 1755bp

[SEQ ID NO:] 3864970 Strep Assembly -- Assembly
 id#3864970
 TTGAGTTAGTACCAATGGACCGACAATTAAAAAGTCATGTTGCTGATTTTCAGAAAAT
 CCTTATCCAGAAATGGAAGAGCAGATGAGGCTGATTGACCGAGTGTGGCCTGAACATTAT
 TTTAAGAACTTAACTCAAGCAACATTAGTCCTGAAACGAATAAAAAATCTGGGAATTA
 ATGCAAGAAAAGGCTTAGAGTTGGAAAATCAAGAATCCAGGAATTTCAGGATATCTGGG
 AGAGATTACTGAGGAAGATTTGAGAATTGTCGGATAGAATCTCATGTCCCTGTATTTA
 TTTTTGTCAGACTTATAGAGAAAAAGAGTACAGAGAATCAGAATATTGGACTTCCAATA
 CTAAACTCATTTAGGAAGGAATCACCATTATTACAATGGTCAGAATCGGAAAAAATTG
 CGGCTATTATTCGAGAATTGTCAGAATAAGATGGAAAAAGGAGATTACAGGAGACAAGA

TGAACACTTTAATGTTGGAAAATCGTTAACGCAGGGATTACAGGGTGAGATGCGAG
 TCTTGTCTGTGACGGATTTGCAGAAGAACGGTTAAAAAAGGAGCTGAGCTGGCTTGT
 TTGATGAAAAAAGATCAGTTGTCCAAACAGTGACCATCGCTAGCCACCGTAAACAGAAGA
 ACTTTGACATTATTAATTCAAAGATATGTACCATATCAACTATCGAAAAGTACAAGG
 GATACAGTCTCAAGGTCGCTGAGGAAGATTGAATGACCTAGACGATGGTGAATTCTACT
 ATCACGAGATTATCGGTTGGAAGTCTATGAGGGTGATAGCTTGGTGGAACCATCAAGG
 AAATCCTGCAACCAGGTGCTAATGATGTCTGGGTGGTCAAACGAAAAGGCAAACGTGATT
 TGCTTTACCTTATATCCCACCAGTGGTCTCAATGTTGATATTCAAATAAACGGTCG
 ATGTGGAATCTTAGAAGGGTTAGACGATGAAGATTGATATTAAACCCTTTCCAGAG
 ATGTTTCTCCACTGGAGCACTCAATCGTGGAAAGGCTCGAGAAAAAGGGCTCTGGAT
 ATCCAGTATCATAATTTGAAAAAAATGCTGAAAAGGCCGTCAAGTTAGATGATGAACC
 CTACAGAGGCGGTCAAGGCATGTTGATCAGAGCACACCTATTATCGAATTCTTAGATG
 CTATTGAAAAGAAAAATCCGCGCGATATTCTCCTCGATCCTGATGGAAAGCAGTTGATC
 AGGCTTATGCTGAAGATTGGCTCAAGAGGAAGAGCTAATCTTATCTGTGGGCACTTAT
 GAGGGTTATGATGAGCGCATTAAGACCTTGGTAACAGATGAGATTCCCTAGGCGACTAT
 GTCCTCACTGGTGGAGAATTGGCAGCTATGACCATGATTGATGCTACAGTCGCCTGATT
 CCAGAAGTGATTGGCAAGGAGTCTAGCCACCAAGATGATAGTTTTCTTCAGGTCTTTA
 GAATATCCTCAGTACACACGTCCTATGATTATCGAGGCATGGTGTGCCAGATGTATTG
 ATGAGTGGCCACCATGAAAAGATTGTCAGTGGCGATTGTACGAGAGTTAAAGAAAACC
 TACGAGCGCAGACCAAGATTACTGAACATTATCAACTGACAGTAGAAGAAGAAAAATG
 CTGGCAGAAATCAAAGGAAACAAAGAATAAAGGAGAACCTATGCAAGTAATCAAACGTA
 ATGGCGAAATTGCGAT

ORF Predictions:

ORF #	Start	End	Direction	Length
7	1309	1710	F	134 aa

[SEQ ID NO: 3864970-7] 3864970-7 ORF translation from 1309-1710, direction F
 VGYEYDERIKTLVTDEISLGDYVLGGELAAMTMIDATVRLIPEVIGKESSHQDDSF
 SGLLEYPQYTRPYDYRGMVVPDVLMMSGHHEKIRQWRLYESLKKTYERRPDLLHYQLTVE
 EEKMLAEIKGNKE*

Blastp and/or MPSearch Result:

Description:

tRNA (guanine-N1)-methyltransferase (trmD) homolog -
 Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3865012
 Assembly Length: 1130bp

[SEQ ID NO:] 3865012 Strep Assembly -- Assembly
 id#3865012
 ATCGAATTCCATAAAATCTTTCTTCCAGATACCCAGACAGGCAATCTCTGGAAAGTT
 CAACGGCCTTATCCGTCTCGCACACAACCATAACATCTCAGAAAAAGCTCTCTCAG
 CCATTTTCAATATCTGCTACGATTGTTCTGGCATAGGGAGGGTCTAAGAAAACGA
 GGTCAAATTCCCCAGATAACCTGTTCCAATGCCCTTCTGCATCCATTTGGAGGAGTTG
 AAATTTCCAATTCCTGGTCATCTGGATATTCAGCCACGATGGTCTGAGCCTTACG
 GTCTCGCTCCACCAAAACAGCACTGGACATGCCACGCATACTGCTTCGATAGATAAAC
 ACCACTACCTGCATAAAGGTCCAAGACTCGTCCCACCTCAAAGTAGGGACCAATCATGTT
 AAAAATGGCTCCCTAACCTATCCGAAGTAGGTCTTGTCTGCCTTAGTGTCTT
 GAGGGGACGTCCCCATAGATTCGTATCGATTTCATACTGTTATTATAACAAATT
 TAGACAAAAAGAGAAAGAAAACGAACCTTGCGGTTGATTCTCTACAAAATATTTCT
 AAGTATCGCGGACTTCTTGAGGCCAACACTTGTGTTGCACTCTCCGATGTGTCTTGC
 GAAGTAGGAACATGCCATACGAGATTGTCATTCCTCCACCGATTGTCAATGGGAATA
 GCCATTCAACAAAGACTTGTGCCATTCAATTCTAACGGTCTTCATCACCTGTAATT
 CCACCTGACGTCTAACAGAGTTCTCATCTACACGAATTCCATAGAAGACAACTCAAAG
 CTCCACCTAACAGACTCATTCCAGACAAGAATATCACCATTAGACCTGTAGCCATTCT
 CAGACTCGCTTGTCCAGTCATCATAGTCTGGTGCACGTCCATCGTGGTTACCATCTT
 GGCAACTCGCCACCGATAACCAATCAAAAGACGGCTCAAATTCTTACAAATCGCATTT
 TCCACGTTCTTAGGTGTCAAGTCTGGTAGCGTTACCAATTCTGTATGGATAAA
 GGTGATTGTTGGCAAGATAGACTCGATGTATAGCGGGCTCAACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	584	973	R	130 aa

[SEQ ID NO:] 3865012-7 ORF translation from 584-973,
 direction R
 VASCDGKPHDGRAPDYDDWTSESENGYKGLNGDILVWNESLGGAFELSSMGIRVDEETL
 RRQVEITGDEDRLLEWHKSLLNGLFPLTIGGGIGQSRMAMFLLRKRHIGEVQTSVWPQE
 VRDTYENIL*

Blastp and/or MPSearch Result:

Description:

asparagine synthetase A (asnA) homolog - *Haemophilus influenzae* (strain Rd KW20)

Assembly ID: 3865148

Assembly Length: 1825bp

[SEQ ID NO:] 3865148 Strep Assembly -- Assembly
id#3865148
TATAACCACCAGGCTCATGACTATAGTCTTTATTCTCTGTAAAAGACTGGTCTTGCA
GATGGCGGTGCAGGCCAAGTGGCTTCGATATAACCCATGATTCTCCTCTTTCA
CAACCAGAAAAGAGGTCTGAATTCTCTCAAATGTGCTTCAAAGACAGAAGGAGGAATGG
CTTCTTCGACCGAAAAATTATCAAATTCAAGTTCAACAATCCGATCCAATCTTCTAATC
TTGCTTGTCTGATTTCATTGTTCTCCAGATAAAAGGGATTAACCAAATCATACTATA
GCCCTGGCTAGTTACATAGAGCAAAGTTCTTCAACACAAACCGTTCAATTCAA
ATAGGAAAGCAGCTCATCAGGACTCTCAAACGAATCCCTTGTAAATCCAGCTCAACTGC
CACCTCTTCAAGGCTGCAAGAAGAAGTGTCCCAGGCCGTCTGTATGGTCAGACTC
GATGACTAAAGAATGTACTTTAGACATTGCGGATTGTCTGACTGGGGACTTGATAAAAT
ATAGCCTAAAAGTTGATTTCATCCCTAGCTAGAAGAAAGGTATCCGACACTTACGGAT
ACTTTCTTCTAAAATATGGGAAAGTTGCTGCTTTCAGCTGGAAAAGACGAGGTCTGAAG
TGCCCCTATCTCAGGCAAATCAAACCTGCTTGCCTGAATGATCTTAATTGGAATTCCAT
GGGAAACATCCTATTGAACATTGCTTGTCAAGTTAGACAAGAGACGCTCAAATGAGTATT
CATAGGTTGGATGTCTCCTGCTCCATAAAGACGTAAACAGCATTGTCATGGTCTAGGA
GTGGAGAACATTTCAACAGTAATCACTGGTGTGTTGATTTATTGGCTAGGT
CTTCTACCTAACGTACCAGTACTTCACGAGCCAGCCATAAATTGCGCTAGAT
AAACAGCATCTGCTGGTTAAAGCATGGCAAAGTCGTCAACAGGGCAATGGTTCTG
TAAAGGTATGCGGTGAAAGAAACTGCTACAATTCTGCTGGATGATGGCAAAGTCATCAATAACTGTA
GCATCCAAGGTGCAATAATTCTGTTGGATGATGGCAAAGTCATCAATAACTGTA
TCATTGACAATTTCAGTGAACAGCTTAAACACGGCAAATGTTCAAGTGCTCA
CGCACCAAGTCAAATCAAACCTGCTGTAAAGAAGACCAATAACGGCTGTCGCAATT
ATGATATTGTGACGACCAAGGTTGGAATGTGGAATTGCCCAAGTTGTCCACGGAAA
TGAACGGTGAAGGTTGAACCAGTTGTTGAACGAAGAAGATCACTAGCTACAAAGTCATTG
CCTTCAGCTCAAAACCATAATAATAATTGGTGCATCAGACGTAATCTTACGCAATTCA
GCATCTCACCATAGACAAAAAGACCCATCGTAATTGTTGGCATAGCGTAAAGGCA
TTGAAAACATCCTCGAGACTTGTGAAATAATCTGGATGGTCAAAGTCATGTTGGTGATA

ATAGAGTATTCTGGGTGGTAAGGCATGAAGTGACGCTCATATCGTCAGATTCAAAGACA
 AAATATTTGGCATTGGCGAACACGACCTGTCCATCTCAAATCAAGAAGCTGGTATCT
 GTAATGTGAGACAAGACATGAGACAAACATACCTGTGTTGAAGTTTCCATGTGCTCCT
 GCTACTCCCAGCTAACAAAGTCACGCATAAAGCTACCTAGAAACTCATGGTAACGTTG
 TAGCTGATACCATTTGGTCCGCAT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	256	423	R	56 aa
7	731	868	R	46 aa

[SEQ ID NO:] 3865148-6 ORF translation from 256-423,
 direction R
 VAVELDYKGIRLESPDELLSYFEMNGFVDEEETLLYVTSQGYSMIWFNPFYLEEQ*

Blastp and/or MPSearch Result:

Description:
 unknown

[SEQ ID NO:] 3865148-7 ORF translation from 731-868,
 direction R
 VITVENVSPLLDHDNAVYVFMGAGDIQTYEYSFERLLSNLTSNVQ*

Blastp and/or MPSearch Result:

Description:
 UDP-N-ACETYLMURAMATE--ALANINE LIGASE (EC 6.3.2.8) (UDP-N-
 ACETYLMURANOYL-L-ALANINE SYNTHETASE) (FRAGMENT). -
 BACILLUS SUBTILIS.

Assembly ID: 3865178
 Assembly Length: 1002bp

[SEQ ID NO: 3865178] 3865178 Strep Assembly -- Assembly
 id#3865178
 ATCGAATTAAAGTAAAACCTAAAGGACTTAGTCCTGTGCAGTACAGAACTAAATCCTTCG
 GATAGAATTATTTGTCTAACCTTTGGGGTCAGTACACCTAAAACCTTGTGATGATATACGT
 TTCCTTGTGAGAATATTTACTTCATTTCCTAAAACCTCAATGTTACTCAGTATTGG
 ATTATGAAAAATCGAGGTCTAAATCTAGATACATTTCTGAAGACAAATCATTGAC
 CACCGAGCAAGAGATTTCAAAAAAGCTGTTAAAACCTCAGAACGTCGCTGAAATCT
 TTGCAATTATCTAACCAAGGCATCAGAAAATATTGGAATGTTGCTGAAATGGTGTAT
 TATCAATATCAAAACCAAACGAAACTCACGAAGATACTGAATCAAAAGACCGTTACTGTCGAG
 TGTTTCCTCGCGAAATGGATGAATCTGCCAGATTCTGAAATAAAAGCTGGATTGTT
 TAACCACATCCGCCTGAGTTAGTGTGCGATATGCAACTTGTCTGATTAACAT
 AATCTAAGGTCAATTGAATCATGGAGTAATCAGAGTACACAACACTTCACCATTCAAAA
 CAGGTTCATTCTTGTGATATTGGTCTGACGAAATTGATCCACCGGAAATAGAGGGTTC
 AAATATATCTGAAACAACTCCTTATGAATAGCAAGTAAGGTGCGAGGACTAAAGCTAAA
 GCCTCTCGAGACAATAGTTCTACAATACGTTAGAGAAACCAAGTCTGCCTCCTCCGT
 ACTTGCATCAATAATATGGTGAATAAGCCGGTGCATTCTCATAAACCTGCTCATAAGTC
 AGTTCTCCCCGGGACTGTTCTCAGCAAAGATTCCATATACGCTGATGGCACTAGATTG
 TCAACTTCTGCAGACCAAAACCTATCCGCCATAATCACGCTTCGCTTCATAAGACAAG
 TTTGGATTGTCAATGTTGTAAGTTGGTGCATAAAAATATCC

ORF Predictions:

ORF #	Start	End	Direction	Length
6	182	580	R	133 aa

[SEQ ID NO: 3865178-6] 3865178-6 ORF translation from 182-580,
 direction R
 VYSDYSMIQM TLDYDFNQE KQVAYATLTQADVVKQI QHFISGIWQIHPFREGNTRTVF
 LIQYLREFGFIDNTPFQQHSKYFRDALVLDNAKILQRRSEFLTAFFENLLLGGQNDLSS
 EKMYLDLDDFS*

Blastp and/or MPSearch Result:

Description:
 unknown

Assembly ID: 3865260
 Assembly Length: 1250bp

[SEQ ID NO:] 3865260 Strep Assembly -- Assembly
 id#3865260
 CTGTCACNACTCCATTTACTACCGATTGCCATGAACACCAACCACACAAAAATGATAT
 AAAGAATGCAATTCCAATAGCACCATAACAAAGATCCAGTTAACCTTGCACCGAACCTG
 AATAGCAGAATAATCATTCTATGAATGTTCCGCCATTAGTCATGACTTCGCTAAAAT
 ATATACAATCATAGAAGATAAGAAAATTACAAATGCTGGAATCATTGCTTCAAACGT
 GGCAATAGCTTGTGGAACCTGTTCTGGCATCTTAATAACAATTTCCTCTTATAAAGAA
 GGTATAAATACTTCCTACTACCAAACCTATAATGATAGCACCATAATTCCCTGGCCTC
 CAAACCAAACTTACTAATAGCGTCCCCAATCGCCTCACCTGTTAGGGATATAAGATG
 ATCTTAGAAAATAAAGAATGCAGATAACAGATAGAACTCCAGCTGGTAAAGCCTCTACTC
 CGCTATTCTTAGCATAAGAATAGGCAATTGAAAAACAAGAAATTAGACCCATAATAGCAA
 AAGTTCCCTGAATATACTTGATAAACGGCTCTGCCAATTAGCTCCAAAACACTAGCAA
 TGCTCTTATTTAATCCTCGAACGGCAATTGTCCTGACATCTTGTGACATAATAATCTCC
 CTGTCAATGGCAAATTGCTAACATCCCCTTTAGAGCTATAATGCCACGCATATTCA
 CAAACTTCATCATCGGTGCAATGATTTCTGAACATCCATCTTGTGACATAATAATCTCC
 TTTCTTACCCACTAATCAAAGATAGGCCAAATCTAATAACTTTCTCATCTAACATA
 CCATAGTCCATCATCGGAATAACAGCTACGGAACATCACACTTATCACAAATTCTTT
 GATTTATCTAATGTATAAGCAACTTGTGGACCCAATAGTGCAACATCTATATTGGCGCA
 TAATCCGCTAATTAGACTGAGAAAACGCCTCTATTCTGCCTCAACTCCACTAGCTTGC
 GCTGCAATTTCATATTATTACAAGCATACCAGTAGAAAAAACCTGCTGCACAAAACAA
 ACCAATCTCACCATTATGTTCTCTATGTTAATAACAATGATAATACTCTAGTA
 ATAATTCTATGAAGTTCTTCTCAAACATAATTCCCTTGAATTAAATTAAATC
 TCCGGTCATACTAGTCCATGAAAANGATCTTGTGAATGAACCAAGAAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	19	399	R	127 aa
7	272	793	R	174 aa
8	786	1073	R	96 aa

[SEQ ID NO:] 3865260-6 ORF translation from 19-399,
 direction R
 VRRLGTLLVKFGLEAKGIIGAIIGLVVGSIYTFIIRKIVIKMPEQVPQAIAKQFEAMI

PAFVIFLSSMIVYILAKSLTNGGTIFIEMIYSAIQVPLQGLTGSLYGAIGIAFFISFLWWF
GVHGNR*

Blastp and/or MPSearch Result:

Description:

cellobiose phosphotransferase system celB - *Bacillus stearothermophilus*

[SEQ ID NO:] 3865260-7 ORF translation from 272-793,
direction R
VGKKRRFIMSKMDVQKIIAPMMKFVNMRGIIALKDGLAILPLTVVGSFLIMGQLPFEG
LNKSIASVFGANWTEPFMQVYSGTFAIMGLISCFSIAYSYAKNSGVEALPAGVLSVSAFF
ILLRSSYIPKQGEAIGDAISKVWFGQGNYRCYHYRFGSRKYLYLLYKEKNCY*

Blastp and/or MPSearch Result:

Description:

cellobiose phosphotransferase system celB - *Bacillus stearothermophilus*

[SEQ ID NO:] 3865260-8 ORF translation from 786-1073,
direction R
VQQVFTGMLVNNMKIAAQASGVEAEIEAFSQSKLADYAPNIDVALLGPOVAYTLDKSKE
ICDKCDVPIAVIPMMDYGMLDGKKVLDLALSLISG*

Blastp and/or MPSearch Result:

Description:

cellobiose phosphotransferase system celA - *Bacillus stearothermophilus*

Assembly ID: 3865272

Assembly Length: 1164bp

[SEQ ID NO: 3865272 Strep Assembly -- Assembly
id#3865272
AATGTAATCGGGCGAGCAAGGACGTGAAGACGCCTTGTAGATCCACTTGCAGATATTGA
TACAATTAATCTGGAATTAAATTCTTGCTGACTTAAAGATCAGTGAACAAACGATATGCGCG
TGTAGAAAAGATGGCACGTACGCAAAAGATAAAGAATCAGTAGCAGAATTCAATGTTTC
TTCAAAAGATTAAACCAGTCCTAGAAGACGGGAAATCAGCTCGTACCATTGAATTACAG
ATGAGGAACAAAAGGTTGTCAAAGGTCTTCTCTTGTAGCAGACTAAACCAGTTCTTATG
TAGCTAATGTGGACGAGGATGTGGTTTCAGAACCTGACTCTATCGACTATGTCAAACAAA
TTCGTGAATTTCAGCGACAGAAAATGCTGAAGTAGTCGTTATTCAGCGCGTGCTGAGG
AAGAAATTCTGAATTGGATGATGAAGATAAAAAAGAGTTCTGAAGCCATTGGTTGA
CAGAATCAGGTGTAGATAAGTTGACCGTGCAGCTTACCACTTGCTGGATTGGAACTT
ACTTCACAGCTGGTAAAAAGAAGTTCGCGCTGGACTTCAAACGTGGTATGAAGGCTC
CTCAAGCAGCTGGTATTATCCACTCAGACTTGAAGAAAGGCCGTAAGAAGCTGGACGCT
TGTCAATGAAGATCTAGTCAAATACGGATCTGAAAAGGCCGTAAGAAGCTGGACGCT
TGCCTGAAGAAGGAAAAGAATATATCCTCAAGATGGCATACTGGAAATTCCGCTTTA
ATGTCTAAAATTAAATAATGGTGTCAATTAGGTTGGAAAAAAATCCAACCCCTTGGC
TTTGAAAGGAAAATAATGACCAAATTACTTGTAGGCTTGGAAATCCAGGGGATAAA
TATTTGAAACAAAACACAATGTTGGTTATGTTGATTGATCAACTAGCGAAGAACAG
AATGTCACTTTACACACGATAAGATATTCAAGAATTGGACCTAGCATCCTTTCC
AAATGGAGAAAAATTATCTGGTAAACCAACGACCTTATGAATGAAAGTGGAAAAGC
AGTTCATGCTTATTAACTTACTATGGTTGGATATTGACGATTACTTATCATTACGA
TGATCTTGACATGGAAGTTGGAA

ORF Predictions:

ORF #	Start	End	Direction	Length
6	101	193	F	31 aa

[SEQ ID NO: 3865272-6 ORF translation from 101-193,
direction F
VNKRYARVEKMARTQDKESVAEFNVSSKD*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3865280
 Assembly Length: 1320bp

[SEQ ID NO:] 3865280 Strep Assembly -- Assembly
 id#3865280
 CGAATTCAAGTTCTTTGTTGTCCTCCATTGTTACGTTAATCTTGAATCGAGG
 GATGATGTTCTTCGAAGCAATTAGTTAGAATCATCTACTGAGGTTATTAATCTGTA
 GAGGTAGAGAGTTGAGTTGAAACAGGAAGACAATATTTCTATCCGGAAAAGAACAA
 GATTGTATTAAGGAATGGCGAATTTCGGTTATTATCTACGAATTGGGACCACCTGT
 TTATCCAATTCTTATTCTTAGGAATGGAATTCCAATGTCGTGAAAACAAGGTAGATGG
 TAGACACTATGTATCAAGATATTACTGGAACTGTTGAAATCACCAAAAAAGTTG
 TGGCTTGATTATTGGGGAGCATGTTCTATAAAAAGAAGAGATTCAAGAGGCATT
 TTTGAATATGTTGAAGGAATAGCTAACCTAGTTATTCCGTAACAGTATAATTCTGG
 TATGACCATATGACCGATATTACAGAGGAAGGTATTTAAAAAGTTCTGAGATTGCA
 GATGGATTGAAAATCATGGAGTTCAATTAGCTTATGCTATGTTGATGATGGTGGACA
 AACTATCAATCAGTTGGAAATTCAATCATAAATTCCAAATGGTTGAGAAATATTAAA
 TATCTTGAAATGGATTGGTCCAACCTAGGATTGTTGGATTGGTCCCCGAGGTGGTTA
 TAATGGACAGAAATCATTATGAGTTGATTGGTTAGAAGCACATCCCAGAGTTAAATAT
 TGGATCTAAAATGGATTCAATGATGTAACGTGGCTGATTAACTATCTCAATCA
 AATGAAGAAAAGATGTTGAAATATCAAAAAGAATTCGATATCAGCTATTGAAAATTGA
 TGGTTGGTTACTTCAACCTGACAAACCTGATAAGAGTGGACCGCACGGTATGTACCAT
 GACAGCGGTTATGAGTTCTTAATTCAACTGTTGATAGATCTAAGAAAGGAGAGAGGAGG
 AAAAGATTGTTGGTTAAACTTGACTTCTATGTAACCTAGTCCATGGTTTACAGTG
 GGTCAATAGTTATGGATTCAAATATCTCAAGATGTAGGCTTACAGAGAATGCAGGTAA
 TGATATCAATCGTATGATTACTTACCGAGATAGTCAGTATCAAGAATTGGAAAAC
 GTGAGATACAGTTACCTATGTTGGTCGCTTTATAATCATGAACCAATCCTATGCTGT
 CAGTGCCAATACCTGGTACATGGATCATCAAATGTTGCATCAATACCAGATTGAAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
7	815	1204	F	130 aa

[SEQ ID NO:] 3865280-7 ORF translation from 815-1204,
 direction F
 VADFNLYLNQMKKKMLEYQKEFDISYWKIDGWLLQPDKPDKSGPHGMYTMTAVYEFLIQLL
 IDLRKERGGKDCWLNLTSYVNPSPWFLQWVNSLWIQISQDVGFENAGNDINRMITYRDS

QYQEFLGKT*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3865286

Assembly Length: 1305bp

[SEQ ID NO:] 3865286 Strep Assembly -- Assembly
id#3865286

CTTAGAAGAAAAGGCTGAGGGCAAATACTAGTCTGTCGAGTTCTTCATTGCGCG
TGATCTCTTCTGGAAAATCTGAAAATCTGGGACGAGAACTGGGTTATCAGCTTCCAAG
TGGAGCTGGAACGGCTCTGACAAGGTGGCTAGCCAGATTGCAAGCCTATGGTATGCA
GGGACTCAACTCTGCGCCAATTGCACTTTAAAAACACTGAAAAAGCGAAAAACGCTT
AGAAAGGTAAGTTATGAATTCAATTAAAAATTCTTAAAGAGTGGGGACTGTTCTCCT
AATTCTGTCAATTACTAGCTTAACTGATCTTTGGAGCAATGTTCGCGTAGAAGG
ACATTCCATGGATCCGACCCCTAGCGGATGGCGAAATTCTCTCGTTGAAAACACCTTCC
TATTGACCGTTTGATATCGTGGTGGCCATGAGGAAGATGGCAATAAGGACATCGTCAA
GCGCGTGTGGAAATGCCTGGCGACACCATTGTTACGAAAATGATAAACTCTACATCAA
TGACAAAGAACGGACGAGCCTATCTAGCAGACTATATCAAACGCTCAAGGATGACAA
ACTCCAAAGCACTTACTCAGGCAAGGGCTTGAAGGAATAAAGGAACCTTCTTAAAG
TATCGCTAAAAAGCCAAGCCTTCACAGTTGATGTCAACTACAACACCAACTTAGCTT
TACTGTTCCCAGAAGGAGAAACCTTCTCGGAGATGACCGCTGGTTTCGAGCGACA
GCCGCCACGTTAGGTACCTTCAAAGCAAAAGATATCACAGGGGAAGCTAAATCCGCTTC
TGGCCAATACCCGTATCGAACATTAAAGAAACCTAAGAGGCCGAGAATACCAATCT
CAGCCTTCTTCTATCGTGAGAAAATGATTGGTACTATCTAAACTTACCAACAGAAA
CACCTCAACTCTCACCTATTGCAAAGGAATTGATGAAAGTTATTTCAGGAACCT
ATTGAACGGATTATTTGAAAATCCAGCAATTGATTCGATCCTCCTCCTAGAAATC
GACGATACGGACGCAGAGGATTTGATGATTTGAAATCATTGTCACAGGAACCATGGCT
GATGTAATTGAGGGCGAAGACTATACTTTGGGGCAAATTGTCCAGCACTCCAAGTAT
GGAGAACAACTGCAAATCAGTCGTTATGATCGCGCAAACCAACTAGTAAGGGCTGGTC
AAGTACTTTCAAGTAGCCATTCAAGGGATTGGTCTCAAGACAG

ORF Predictions:

ORF #	Start	End	Direction	Length
-----	-----	-----	-----	-----
6	146	250	F	35 aa

[SEQ ID NO:] 3865286-6 ORF translation from 146-250,
direction F
VASQILQAYGMQGLNFCAKLHFKNTEKAKKRLER*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3865326
Assembly Length: 804bp

[SEQ ID NO:] 3865326 Strep Assembly -- Assembly
id#3865326
CTATGCTTGTAAAGGGCTTGCTTCAGGATCAGTTGCCTTACGGTCACTGGCATTCCAACCT
TGGTCTTGAAGGGGGACTATCTTGCCTGAGCAACTCTGGGTGTTATCGAAATTATCCGT
ATCTTATCATCAATGGTGGAAAGTCTTACAAATGGTGCAGGTATCTTAAGGATTCT
AACTTACAACCTGGCAAATGGTTACTTCTTGTGATTACAACCATTGCAACCTTG
AACTTCTTGCCTGAGCCAATTGGACGTTCAACCCCTCTCTGTTGAGATGAAATCGCT
GCTGAGTCAGTTGGGTTAATACGACTAAAATTACGTTTGTCTTGCGCC
ATTACTGCAAGTATTGCTGGGTCACTTCAGCCAGGATTAATCGGCTGTGTTGACCGAAA
GATTACACCTTCATCAACTCAACGTTTGATTATTGTTGATTGGTGGACTCGGT
TCCATTACAGGTGCGATTGTTGGCTATTGTTCATCGAATTGGATATGCTTCTCCAA
GATGTTGCTAGTGTGGTATGATTATTACGCTTGGCCTGGTATTGTAATGATTTC
AGACCAAGGTGGACTCCTGGAACGTGGAACTGAGCCTATCACGTTCTTAAAAAATCT
AAGAAGGAGGAACAAAATACTGGCATTACTGAAGTAAAACAGTTAACCAAACATTG
GTGGTCTAACAGCTGGAGATGTGACTCTGGAATTGAACGAAGGGAACTGGTTGGAT
TAATCGGTCCAAACGGAGCTGGGA

ORF Predictions:

ORF #	Start	End	Direction	Length
				159

7 100 681 F 194 aa

[SEQ ID NO: 3865326-7] 3865326-7 ORF translation from 100-681, direction F
VFIEIIRIFIINGSLTNGAAGILRIPNFTTWQMVFVVIITIATLNFLRSPIGRSTLS
VREDEIAAESGVNTTKIKIIAFVFGAITASIAGSLQPGLIGSVVPKDYTFINSINVLI
VVFGGLGSITGAIVSAIVHRILNMLLQDVASVRMIIYALALVLVMIFRPGGLGTWELSL
SRFFKKSKKEEQN*

Blastp and/or MPSearch Result:

Description:

HIGH-AFFINITY BRANCHED-CHAIN AMINO ACID TRANSPORT PROTEIN
BRAE. - PSEUDOMONAS AERUGINOSA.

Assembly ID: 3865438

Assembly Length: 553bp

[SEQ ID NO: 3865438] 3865438 Strep Assembly -- Assembly
id#3865438
CCCATCTGCCTTGACCAAAGGCTACCACTTCAAAACTCGCCTCACCCGGAAATTCA
GCTTAGATGGCATTACCTGCCCGAGTAGTACGAGCCTTCGACCTGAAAATTCTG
ATATAAAAATAGGTTCTGATTATCCATTCAAAGGAGCTAACGTTCAAAACTTTG
ACCGTTCCAAGCTAAGTGCCTCCAAATCCAACCTTCATCTAGGTTAACATTCTT
CCACCAGCATCTGCACCTTTCACGAACATAATCTCCAAACCTGAGATAATCTGAG
AGTTGCTCAACTTCCAGCGTCATACCCGCTGCACCTGCATGACCTCCAAAGGCGATGAAG
AGGTCTCGATGGGATCCAGAGCTTAAAAATATCGACCGCTTCCACACTACGAGCACTG
CCCTTGGCACGACCGTCTCTATATTAAGAAACAATGACTGTCTGCCCATTCTCAA
TAAACGACCAGCCACGATTCTAGAACCCAGGATTCCAGCCTCCTGGCCAAGACCTG
AACTTTCTCAG

ORF Predictions:

ORF #	Start	End	Direction	Length
160				

6

75

407

R

111 aa

[SEQ ID NO: 3865438-6] 3865438-6 ORF translation from 75-407,
direction R
VEAVDIFEALDPHRDLFIAFGGHAGAAGMTLEVEQLSDLSQLVEDYVREKGADAGGKNKL
NLDEELDLEALSLETVKSFERLAPFGMDNQKPIFYIKNFQVESARTTGGR*

Blastp and/or MPSearch Result:

Description:

unknown

Assembly ID: 3865446

Assembly Length: 965bp

[SEQ ID NO: 3865446] 3865446 Strep Assembly -- Assembly
id#3865446
ACATCTTAAGATTAATTCAGAATCTCTCTGAAGACTTTAAAGTTGGTCGTCTATA
GGGAGTTTGGCCATCGTTGCTCAATTGTCTGATTAAGGTCTACCCCTGATGAAACAA
TTATTATCCATGTTCTTTATTATAGACAAAGTAAGAAGACGTTCTGAATGTAGACT
TTATATTTTATGATTTCTCTTCCATAATATCCAATTGATAGTTGGAAATGAAAATA
AGACCGCTCTGTTGACACCGAAAGACACCTTGATATAGACGCCCTATCAACTAGCTC
TCTATTGGTCTCTGCAAGTCCACTTCAAATTCAACGGTATCTCATTTCCTTA
AATGTCTAAAGGCTCCTCAATCTCTCAGTGGATACTTTATCCTTATCTCGTTCTTCT
TGGAAAGCATGGTACTGTTCTGTAAATTCTCTAATCCTCTGAAGCAACGACTTCCTTA
TTTTAAAATAATCTGAAAAAATTGACATCATATAATTCTTATCACTTATTTTTGA
TGACCCAAACTTATCTTGATTATTTCTCAGGATAAAAGTTACATTTTTGTTTT
AAGTCATGGTAGATTCAATTCTTGCTTTGTTATTAATCTCTAAAGAATTGACA
CGGTTAACAAAATTCTAACGACTTCAATCTTGCTTAGCAAAATGCCTCTAAAA
AATTCTTCATCATATAGATCTCGTTGCTGAGTTGGCGCCCTCGAATTGGTTTATCATC
GTTCTATCTGTCATAAAAACGGCTATGCTTTGACTAAAATCAATCTGAACATGCAAC
TGCTTGCTTCTCTAAAAATCATCAAACGATTAGATTGCTGAAGCAAAAATAAAGA
CGTGTTCATTCAAATTGACTAGATTCTTATATTTTATAATCTCGATAGGAA
TAACG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	42	326	R	95 aa

[SEQ ID NO:] 3865446-6 ORF translation from 42-326,
direction R
VELAENQIEKLVDKGVYIKVSGVKQSGLIFIPNYQLDIMEEENHKKYKVVYIRETSSYFV
YNKENMDNNCFIKGRTLIRQLSNDGQKLPIDDQL*

Blastp and/or MPSearch Result:

Description:
unknown

Assembly ID: 3865474
Assembly Length: 795bp

[SEQ ID NO:] 3865474 Strep Assembly -- Assembly
id#3865474
TCCCAAGCAAATCCTTGATAGCATGGACTTGCTGTCAACGTTCATGCCTCCTCCTCC
TAGACACCCTGGTGGTGCCTATCCATTATGCCTTGATTCAAGGGGATGAGGAAGCTGG
TGTGACCACATGGAAATGGTTAAGAAAATGGATGCAGGAGATATGATTCTCGTCGCAG
CATTCGATCACAGATGAGGACAATGTTGGCACCTTGTGAAAATTGGCGTAGTTGG
TCGTGATTGCTTTGGACACTCTGCCTGCCTATATTGCTGGTGATATCAAACCTGAACC
GCAGGGATACGGAGTCAGGTTACCTCTCCAAATATAAGCCAGAGGAAGAAAAACTGG
ACTGGAACAAACCAATCGTCAACTCTTAAACCAATTCTGGAATGAACCCCTGGCCTG
TTGCCCATACTTCCTTAAGGGCGACCGCTTAAGATTATGAAGCCTACCACTAGAAG
GTCAGGGAAATCCAGGTGAAATTCTCTATCGGCAAGAAAGAATTGATTGTCGCAACGG
CTGAAGGGCTCTATCCCTCAAACAAGTGCAGCCAGCTGGTAAGCCTAACGATGGACATTG
CTTCCTCCTCAACGGAGTTGGACGTACATTGACTGTAGGAGAACGATTGGTACTAAA
GTAGAAACGGCTAGAAGTTAGCTCTAGCAGTGCTAGAGGATGTTTGTGAACCAAGCA
TATTCAAATATCGCCTTAAATAAACACCTCAAGGGAGTCAGCTTCTGCAGCAGACAAG
GGCTTAGTGACCGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	243	659	F	139 aa

[SEQ ID NO: 3865474-6] 3865474-6 ORF translation from 243-659, direction F

VICFWTLCLPILLVISNLNRRIRSQVTFSPNIKPEEEKLDWNKTNRQLFNQIRGMNPWPV
AHTFLKGDRFKIYEALPVEGQGNPGEILSIGKELIVATAEGALSLKQVQPAGKPKMDIA
SFLNGVGRTLTVGERFGD*

Blastp and/or MPSearch Result:

Description:

methionyl-tRNA formyltransferase (fmt) homolog - Haemophilus influenzae (strain Rd KW20)

Assembly ID: 3865476

Assembly Length: 816bp

[SEQ ID NO: 3865476] 3865476 Strep Assembly -- Assembly
id#3865476

CTGGTAAAATTGAGGAAACCTTGTATGGTCTAAAAGACAAGTACACCATGCTCTGGTAA
CCCGTNCCATGCAGCAAGCTTCACGTATCTCTGATAAGACAGGATTTCTAGATGGAG
ATTTGATTGAATTAAATGATAACCAAGCAGATGTTCTTAATTCCCCAACACAAGGAAACG
GAAGACTATATTACAGGAAATTGGATAAGGAGATGAAAGATGTTACGATCTCAATTG
AAGAAGATTAGAGAAATTACATAACCAGTTACGCTATGGACAAGAAGTGTCTCAC
AAATCAATCCGTACGGTACGTGCTTGTACGCATGACCGTGACCTGGCAAAAGAGGTC
ATCGAAGATGATGCAGAAGTAAATGAATACGAAGTGAAACTGGAAAGAAATCATTGAA
ATGATCGCACTCCAACAACCAGTCTCTCAAGATTGCGTACAGTCTGACTGTCCTTAAG
GCTGTATCAGATGTGGAGCGTATGGGGATCACGCTGTAGCCATTGCTCAGGCAACCATC
CGTATGAAGGGGAAGAGCGCATTCCAGCTGTAGAGGAAGAAATTAAAAGAAATGGGACG
TGAAGTTAAAAGCGTTGTTGAAGCAGCACCTGATCTTATCTTAATGGTTCTGTTGACGA
CGCATAACCGGGTGGCCTCCATGGGATGAGCAAATTAAACCACTATTTGAAACTATCCGTG
AACCTTGCAGTGAATGAAGATTAAGAAGAGTCCAATCCAGAAGCCATTGTGACGGGTC
GTGATTATTCAGTTATTCCTACTTGGGAGCGT

ORF Predictions:

ORF #	Start	End	Direction	Length
6	394	603	F	70 aa

[SEQ ID NO:] 3865476-6 ORF translation from 394-603,
 direction F
 VKLEKKSFEMIALQQPVSQDLRTVLTVLKA VSDVERMGDHAVAIAQATIRMKGEERIPAV
 EEEIKRNGT*

Blastp and/or MPSearch Result:

Description:
 Probable phosphate regulator PhoU homolog

Assembly ID: 3865502
 Assembly Length: 1041bp

[SEQ ID NO:] 3865502 Strep Assembly -- Assembly
 id#3865502
 CTGAAATTGCACCACCAAGATGGGATTGGGCAGGTTCTCAGCAACCTCTTGCTCAAACCTGG
 TTGACAACCCAGTCACGCCCTGCTTACTGCTAACTATATTAGAATCTTATCTTGGGCAG
 TCATTTTGAATCGCTATGAGAGAAGCCAGTAAAAATAGTAAAGAATTGCTAAAAACTA
 TCGCTGACGTGACTTCTAAAATTGTCGAATGGATCATCAATCTGGCTCCATTGGAATCC
 TTGGTCTTGTAAAAACCTTCTGACAAGGGAGTCGGAAGCCTGCCAACTACGGTA
 TTTTATTGGTTCTATTAGTAACGACTATGCTTTGTTGCCCTGTGGTCAACCCATTGA
 TTGCCTTCTCTTTATGAGACGCAATCCTTACCCCTAGTTGGAACTGCCTCCGTGTC
 AGCGGGTGTGACAGCCTTTCACTCGTAGTTCTACGACTAACATTCTGTCAACATGAA
 ACTCTGCCATGACCTTGGACTCAACCCAGATACTTCTGTTACAGTTAA
 TACTATCAATATGGCTGGAGTAGCGATTACCAACCTTTGACCCTGTTACAGTTAA
 CACTCTGGAAATTCCCTGTTGACTTGCACAGCCTTGTCTCAGTGTGGTAGCAGCTAT
 CTCAGCCTGTGGTCTCAGGTATTGCCGGAGGTTCCCTCCTTATCCCAGTTGCTTG
 TAGCCTTTGGTATTCTAACGATATTGCCATACAAATTGTTGGGGTTGGTTTGAT
 TGGTGTCACTTCAAGACTCATGTGAAACAGCCCTTAACCTACAGATGTCTCTTAC
 CGCCGTTGCCGAATACGCAGCAACCCGTAAAAAAACTCATCAAGGCAAGCCTGCTTAT

GTCTTGTCTTTACGCTTTATTCTAACCTATTAGGAAATTCTTATGTCTATTAGCCAAC
 GTACGAACAAGCTCATCTTAGCTACCTGTCTGCCTGCCTGCTTATTCTCAATC
 TTTCATCAGCAGTTCGGCTG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	428	877	F	150 aa

[SEQ ID NO:] 3865502-6 ORF translation from 428-877,
 direction F
 VTAAFFRSSTTNIPVNMKLCHDLGLNPDTYSVSIPLGSTINMAGVAITINLLTVNTL
 GIPVDFATAFVLSVVAAISACGASGIAGGSLLLIPVACSLFGISNDIAIQIVGVGFVIGV
 IQDSCETALNSSTDVLFTAVAEYAACTRKK*

Blastp and/or MPSearch Result:

Description:
 Probable sodium-dicarboxylate symporter

Assembly ID: 3865694
 Assembly Length: 544bp

[SEQ ID NO:] 3865694 Strep Assembly -- Assembly
 id#3865694
 CTGATGACACAAAGCACAGTGGTAGGACTTGCGAAGTCACCCCTTTCTTCAAAATT
 ATACTAAATCATTGATATCAGTGTAGTCACGATTAAGTCCTTGAGCAACTGGTAGGCTAG
 TCAAGTAACCTTGATAAGTGGTCACACCTTGACGCAAGCCTTCATCTTCAGAGATTGCTT
 GTGCGAATCCTTGCCAGCAAAGCTCGATATAAGGAAGAGTGACATTGGTTAGGGCGA
 TGGTTGAAGTGCAGGCAACCGCACCAGGGATATTGGCAACGGCATAGTGGAGAACACCGT
 GTTTTCATAGACGGGTTCATCGTGCCTTGTACACGGTCAGCTGTTCGATAACGCCAC
 CTTGGTCAACAGCAACGTCAACGATAACAGAGCCTGGACGCATTGTTGACCATCTCCTC
 TGTCAACCAATTCCGGTGCTTGCACCAGGGATGAGAATGGCTCCAATCACCACATCAGC
 ATCTCTCATTTGCTCAATGTTGAATTAGATATAAGAATTGAAATTGACTTCC
 AAAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	59	334	R	92 aa

[SEQ ID NO:] 3865694-6 ORF translation from 59-334,
 direction R
 VTTHDEPVYEKHGVLYAVANI PGAVARTSTIALTNVTL PYIEALAGKGFAQAI SEDEGL
 RQGVTTYQGYLTSLPVAQGLNRDYTDINDLV*

Blastp and/or MPSearch Result:

Description:

ALANINE DEHYDROGENASE (EC 1.4.1.1). - BACILLUS SPHAERICUS.

Assembly ID: 3865704

Assembly Length: 810bp

[SEQ ID NO:] 3865704 Strep Assembly -- Assembly
 id#3865704
 CTGCGACTAGCGGATCTCAGACAGAAGGTCAATATGGAAAAGTACATGAAAATGTGATGG
 ACTACTGGTTCAAAACGCATCCAGAAAATTTTCGATAATGTCGGACCTCTGTAGCCA
 GTAACCTTTTCATACTTACACCGAAGATTCCACTTGATGAAGGAAATTGGAGTTAATT
 CTTTCCGACTTCCATCCAATGGAGTCGACTCATCAAGAATTAGAGACAGGTGAGCCTG
 ATCCAAAAGGTATTGCTTCATAATGCCATTGGAAGAAGCTAAAAGAACAGATG
 GATCTTGTGATGAATTACATCATTGATTTACCGAGTGGAACTTCTTCAAAAATACGGT
 GGTGGGAAAGCAAACATGTAGTGGAGTTATTCGTGAAGTTGCCAAGACTGCTTAAACA
 TGCTTGGAGATAAGGTTCAATTACTGGACAATTCAATGAGCCAATGGTCATTCCAGAA
 GCAGGATACTTATATGCTTCATTATCAAATCTAAAGGAAAGGGAAAAGAGGCCGTA
 CAAGTCATCTATAATCTAAACCTGCTAGTGCAAAAGTGAATTCAACTATATCGCTCATTA
 GGACTTGATGGAAAGATTGGGATTATTTAAACTTGACACCTGCTTATCCAAGAAGTAAT
 TCTCCAGAAGACTTACAAGCAAGTCGATTACAGATGACTTCTTAACAAAGTCTCCCTT
 GAATCCAGCTGTTAAAGGAACCTTCCCAGAAAAGATTGGTAAAACAGCTAGAGAGAGAT
 GGCCTGTTATGGAGTCATACCGAAAAAGAG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	232	735	F	168 aa

[SEQ ID NO:] 3865704-6 ORF translation from 232-735, direction F

VSLIQKVLLSTMPFMEEAKKNQMDLVMNLHHFDLPVELLQKYGGWESKHVVELFVKFAKT
ALTCFGDKVHYWTTFNEPMVIPEAGLYAFHYPNLKGKGKEAVQVIYNLNLASAKVIQLY
RSLGLDGKIGIILNLTPAYPRSNSPEDLEASRFTDDFFNKVFLESSC*

Blastp and/or MPSearch Result:

Description:

BETA-GLUCOSIDASE A (EC 3.2.1.21) (GENTIOBIASE) (CELLOBIASE)
(BETA-D- GLUCOSIDE GLUCOHYDROLASE). - CLOSTRIDIUM
THERMOCELLUM.

Assembly ID: 3865788

Assembly Length: 437bp

[SEQ ID NO:] 3865788 Strep Assembly -- Assembly
id#3865788

AATCGCGTATCTCCCTCTTCCCTAACGATTGCTGAAAAATGAGTGGAGGAAAGTTAAT
ACCATTCTCCAGTGTAAATGGTAAATTCCCTTTGAAACATTTTATCATTACTCCTGC
CCGTTGTTACGATATCAGTAGTATAAAATCGACCCCTCTCCCAAAAGAAATTACGTCT
TACATTTTATTTCAATTTCATATAAACTACTCTCACTCAATTGATTACGCTA
TCAATCAAGTCTGGTAATGGATAGGTAAATGTGGAACTTCTCAAACGTGCAAAACAA
ATTCCTTGTAGGCATTGGCGTCCAGCTTCTGAAATTTCACCTCACTCCATCATGA
AGAAAGCTCATTCTTTACGTTTCTTACTAATACCAAGAAGAGCTAAAGGACCTATA
GGTTGTTCAAATACATG

ORF Predictions:

ORF #	Start	End	Direction	Length
6	210	344	R	45 aa

[SEQ ID NO: 3865788-6] ORF translation from 210-344,
direction R
VKISESWTTNAYKGICFAQFGEVPHFTYPLPDLIDSVIKIELRE*

Blastp and/or MPSearch Result:

Description:
unknown

Provided in Table 2 is information on the direction of the ORF (forward or reverse) for each polynucleotide in Table 1. Also listed for each ORF is its start and stop codon positions (refer to the columns containing nucleotide code labeled "Start" and "Stop"). The triplet codon sequence for each start and stop codon is also shown. These codons may be shown in the sense orientation or antisense orientation, such as GTG and CAC, respectively, for start codons. The "Length" column discloses the length of each polynucleotide assembly. The direction of translation on the polynucleotide depicted is denoted by and "Forward" for forward or and "Reverse" for reverse (or being on the opposite strand from the one depicted). As indicated above, the "Assembly ID" number is a unique identifier assigned to each ORF of Table 1 and allows a correlation between the data in Tables 1 and 2.

TABLE 2

Quality Assembly	ORF	Codon	Codon	Position	Position	Length	Direction	
ID	#	Start	Stop	Start	Stop			
Full	3047950	6	~CAC	TCA~	2	451	150	Reverse
Full	3049152	6	~CAC	TCA~	24	407	128	Reverse
Full	3174820	7	GTG	TAG	598	1041	148	Forward
Full	3175500	8	GTG	TAG	714	1049	112	Forward
Full	3175674	6	GTG	TAG	126	314	63	Forward

Quality Assembly	ORF	Codon	Codon	Position	Position	Length	Direction
------------------	-----	-------	-------	----------	----------	--------	-----------

	ID	#	Start	Stop	Start	Stop	
Full	3176442	6	GTG	TGA	350	478	43 Forward
Full	3176630	6	GTG	TAA	273	419	49 Forward
Full	3176662	6	~CAC	TTA~	2	226	75 Reverse
Full	3857692	6	GTG	TAA	386	634	83 Forward
Full	3857944	7	~CAC	TCA~	1332	1475	48 Reverse
Full	3858118	7	~CAC	CTA~	948	1160	71 Reverse
Full	3858152	6	~CAC	TCA~	546	836	97 Reverse
Full	3858258	6	GTG	TAA	207	722	172 Forward
Full	3858314	6	~CAC	TTA~	5	661	219 Reverse
Full	3858368	9	~CAC	TCA~	1207	1578	124 Reverse
Full	3858556	6	GTG	TAA	49	702	218 Forward
Full	3858562	6	~CAC	TTA~	14	178	55 Reverse
Full	3858656	6	GTG	TAA	245	559	105 Forward
Full	3859118	6	GTG	TGA	314	661	116 Forward
Full	3860084	6	~CAC	CTA~	294	473	60 Reverse
Full	3860172	8	~CAC	TCA~	1724	1888	55 Reverse
Full	3860242	7	GTG	TAA	573	1001	143 Forward
Full	3860282	6	GTG	TAA	288	1190	301 Forward
Full	3860296	8	~CAC	TCA~	1697	1843	49 Reverse
Full	3860406	6	GTG	TAA	148	504	119 Forward
Full	3860406	7	GTG	TAA	497	1405	303 Forward
Full	3860416	6	~CAC	TTA~	72	281	70 Reverse
Full	3860712	6	~CAC	CTA~	74	499	142 Reverse
Full	3860728	6	GTG	TAG	259	519	87 Forward
Full	3860794	6	~CAC	TTA~	184	915	244 Reverse
Full	3860830	6	GTG	TGA	176	286	37 Forward
Full	3860984	6	GTG	TAA	113	520	136 Forward
Full	3861088	6	~CAC	TTA~	46	474	143 Reverse
Full	3861138	6	GTG	TAG	42	437	132 Forward
Full	3861256	6	~CAC	TTA~	13	207	65 Reverse
Full	3861256	7	~CAC	TTA~	236	529	98 Reverse
Full	3861262	6	GTG	TGA	181	594	138 Forward
Full	3864150	7	GTG	TAA	922	1998	359 Forward
Full	3864150	8	GTG	TAG	2031	2759	243 Forward
Full	3864190	8	GTG	TAG	1259	1534	92 Forward
Full	3864204	8	~CAC	TTA~	1092	1835	248 Reverse
Full	3864212	6	~CAC	TCA~	256	1155	300 Reverse
Full	3864214	9	~CAC	TCA~	2812	3150	113 Reverse
Full	3864226	8	GTG	TAG	1992	2744	251 Forward
Full	3864242	6	GTG	TAA	376	1002	209 Forward

Quality Assembly	ORF	Codon	Codon	Position	Position	Length	Direction	
	ID	#	Start	Stop	Start	Stop		
Full	3864254	6	~CAC	CTA~	117	833	239	Reverse
Full	3864296	7	~CAC	TTA~	944	1777	278	Reverse
Full	3864296	10	~CAC	TTA~	2323	2694	124	Reverse
Full	3864300	9	GTG	TAA	2479	2823	115	Forward
Full	3864312	7	~CAC	TCA~	736	906	57	Reverse
Full	3864336	6	~CAC	TTA~	295	2232	646	Reverse
Full	3864344	8	~CAC	TTA~	1147	1503	119	Reverse
Full	3864352	6	~CAC	TCA~	303	1808	502	Reverse
Full	3864352	7	~CAC	CTA~	1818	2528	237	Reverse
Full	3864366	7	GTG	TAA	939	1670	244	Forward
Full	3864384	8	~CAC	CTA~	1717	2025	103	Reverse
Full	3864400	7	GTG	TAA	371	937	189	Forward
Full	3864416	7	~CAC	TTA~	929	1189	87	Reverse
Full	3864424	7	~CAC	TCA~	388	1008	207	Reverse
Full	3864430	7	GTG	TGA	627	1100	158	Forward
Full	3864442	7	GTG	TAA	867	1322	152	Forward
Full	3864442	8	GTG	TAA	1562	2074	171	Forward
Full	3864450	7	GTG	TAA	897	1448	184	Forward
Full	3864482	6	~CAC	TCA~	505	1170	222	Reverse
Full	3864496	6	~CAC	TCA~	1	1128	376	Reverse
Full	3864514	6	~CAC	TTA~	551	937	129	Reverse
Full	3864518	8	~CAC	CTA~	1985	2371	129	Reverse
Full	3864522	7	~CAC	TTA~	310	1458	383	Reverse
Full	3864568	6	GTG	TAA	296	493	66	Forward
Full	3864590	6	~CAC	CTA~	125	511	129	Reverse
Full	3864596	11	GTG	TAA	1915	2097	61	Forward
Full	3864624	6	GTG	TAA	446	751	102	Forward
Full	3864630	8	GTG	TAA	663	953	97	Forward
Full	3864654	9	GTG	TAA	1878	2306	143	Forward
Full	3864658	7	~CAC	TTA~	892	1029	46	Reverse
Full	3864664	7	GTG	TAG	675	1727	351	Forward
Full	3864700	6	~CAC	TTA~	480	740	87	Reverse
Full	3864706	6	~CAC	CTA~	336	626	97	Reverse
Full	3864710	6	GTG	TAA	442	972	177	Forward
Full	3864710	7	GTG	TGA	1247	1438	64	Forward
Full	3864724	6	~CAC	TTA~	133	1197	355	Reverse
Full	3864734	7	GTG	TAA	897	1601	235	Forward
Full	3864740	6	~CAC	CTA~	4	264	87	Reverse
Full	3864792	6	~CAC	TTA~	346	1149	268	Reverse
Full	3864830	6	~CAC	CTA~	515	1123	203	Reverse

Full	3864830	7	~CAC	TTA~	1134	1322	63	Reverse
Quality Assembly								
	ORF	Codon	Codon	Position	Position	Length	Direction	
	ID	#	Start	Stop	Start	Stop		
Full	3864848	6	~CAC	TTA~	707	1546	280	Reverse
Full	3864878	6	GTG	TAA	95	622	176	Forward
Full	3864950	6	~CAC	TCA~	198	500	101	Reverse
Full	3864954	6	GTG	TGA	414	1070	219	Forward
Full	3864962	6	~CAC	TTA~	195	602	136	Reverse
Full	3864970	7	GTG	TAA	1309	1710	134	Forward
Full	3865012	7	~CAC	CTA~	584	973	130	Reverse
Full	3865148	6	~CAC	TCA~	256	423	56	Reverse
Full	3865148	7	~CAC	CTA~	731	868	46	Reverse
Full	3865178	6	~CAC	TTA~	182	580	133	Reverse
Full	3865260	6	~CAC	CTA~	19	399	127	Reverse
Full	3865260	7	~CAC	TTA~	272	793	174	Reverse
Full	3865260	8	~CAC	TTA~	786	1073	96	Reverse
Full	3865272	6	GTG	TAA	101	193	31	Forward
Full	3865280	7	GTG	TGA	815	1204	130	Forward
Full	3865286	6	GTG	TAA	146	250	35	Forward
Full	3865326	7	GTG	TAA	100	681	194	Forward
Full	3865438	6	~CAC	TTA~	75	407	111	Reverse
Full	3865446	6	~CAC	TTA~	42	326	95	Reverse
Full	3865474	6	GTG	TAA	243	659	139	Forward
Full	3865476	6	GTG	TGA	394	603	70	Forward
Full	3865502	6	GTG	TAA	428	877	150	Forward
Full	3865694	6	~CAC	TTA~	59	334	92	Reverse
Full	3865704	6	GTG	TAA	232	735	168	Forward
Full	3865788	6	~CAC	CTA~	210	344	45	Reverse

EXAMPLES

The examples below are carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The examples are illustrative, but do not limit the invention.

Example 1**Isolation of DNA coding for a virulence gene in *Streptococcus pneumoniae***

As mentioned above each of the DNAs disclosed herein by virtue of the fact that it includes an intact open reading frame is useful to a greater or lesser extent as a screen for identifying antimicrobial compounds. A useful approach for selecting the preferred DNA sequences for screen development is evaluation by insertion-duplication mutagenesis. This system disclosed by Morrison et al., *J. Bacteriol.* 159:870 (1984), is applied as follows.

Briefly, random fragments of *Streptococcus pneumoniae*, strain 0100993 DNA are generated enzymatically (by restriction endonuclease digestion) or physically (by sonication based shearing) followed by gel fractionation and end repair employing T4 DNA polymerase. It is preferred that the DNA fragments so produced are in the range of 200-400 base pairs, a size sufficient to ensure homologous recombination and to insure a representative library in *E. coli*. The fragments are then inserted into appropriately tagged plasmids as described in Hensel et al., *Science* 269: 400-403(1995). Although a number of plasmids can be used for this purpose, a particularly useful plasmid is pJDC9 described by Pearce et al., *Mol. Microbiol.* 9:1037 (1993) which carries the erm gene facilitating erythromycin selection in either *E. coli* or *S. pneumoniae* previously modified by incorporation of DNA sequence tags into one of the polylinker cloning sites. The tagged plasmids are introduced into the appropriate *S. pneumoniae* strain selected, inter alia, on the basis of serotype and virulence in a murine model of pneumococcal pneumonia.

It is appreciated that a seventeen amino acid competence factor exists (Havastein et al., *Proc. Nat'l. Acad. Sci. USA* 92:11140-44 (1995)) and may be usefully employed in this protocol to increase the transformation frequencies. A proportion of transformants are analysed to verify homologous integration and as a check on stability. Unwanted levels of reversion are minimized because the duplicated regions will be short (200-400 bp), however if significant reversion rates are encountered they may be modulated by maintaining antibiotic selection during the growth of the transformants in culture and/or during growth in the animal.

The *S. pneumoniae* transformants are pooled for inoculation into mice, eg., Swiss and/or C57B1/6. Preliminary experiments are conducted to establish the optimum complexity of the pools and level of inoculum. A particularly useful model has been described by Veber et al. (*J. Antimicrob. Chemother.* 32:432 (1993) in which 10^5 cfu inocula sizes are introduced by mouth to the trachea. Strain differences are observed with respect to onset of disease e.g., 3-4 days for Swiss mice and 8-10 days for C57B1/6.

Infection yields in the lungs approach 10^8 cfu/lung. IP administration is also possible when genes mediating blood stream infection are evaluated. Following optimization of parameters of the infection model, the mutant bank normally comprising several thousand strains is subjected to the virulence test. Mutants with attenuated virulence are identified by hybridization analysis using the labelled tags from the "input" and "recovered" pools as probes as described in Hensel et al., Science 269: 400-403(1995). *S. pneumoniae* DNA is colony blotted or dot blotted, DNA flanking the integrated plasmid is cloned by plasmid rescue in *E. coli* (Morrison et al., J. Bacteriol. 159:870 (1984)) and sequenced. Following sequencing, the DNA is compared to the nucleotide sequences given herein and the appropriate ORF is identified and function confirmed for example by knock-out studies. Expression vectors providing the selected protein are prepared and the protein is configured in an appropriate screen for the identification of anti-microbial agents. Alternatively, genomic DNA libraries are probed with restriction fragments flanking the integrated plasmid to isolate full-length cloned virulence genes whose function can be confirmed by "knock-out" studies or other methods, which are then expressed and incorporated into a screen as described above.

What is claimed is 1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- (a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence of Table 1;
- (b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;
- (c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence of Table 1;
- (d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and
- (e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

- 2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.
- 3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.
- 4. The polynucleotide of Claim 2 comprising the nucleic acid sequence selected from the group consisting of the nucleic acid sequences set forth in Table 1.
- 5. The polynucleotide of Claim 2 which encodes a polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
- 6. A vector comprising the polynucleotide of Claim 1.
- 7. A host cell comprising the vector of Claim 6.
- 8. A process for producing a polypeptide comprising: expressing from the host cell of Claim 7 a polypeptide encoded by said DNA.
- 9. A process for producing a polypeptide or fragment comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide or fragment.
- 10. A polypeptide comprising an amino acid sequence which is at least 70% identical to an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
- 11. A polypeptide comprising an amino acid sequence selected from the group consisting of the amino acid sequences set forth in Table 1.
- 12. An antibody against the polypeptide of claim 10.

13. An antagonist or agonist of the activity or expression of the polypeptide of claim 10.

14. A method for the treatment or prevention of disease of an individual comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 10.

15. A method for the treatment of an individual having need to inhibit a bacterial polypeptide comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 13.

16. A process for diagnosing a disease related to expression or activity of the polypeptide of claim 10 in an individual comprising:

(a) determining a nucleic acid sequence encoding said polypeptide, and/or

(b) analyzing for the presence or amount of said polypeptide in a sample derived from the individual.

17. A method for identifying compounds which interact with and inhibit or activate an activity of the polypeptide of claim 10 comprising:

contacting a composition comprising the polypeptide with the compound to be screened under conditions to permit interaction between the compound and the polypeptide to assess the interaction of a compound, such interaction being associated with a second component capable of providing a detectable signal in response to the interaction of the polypeptide with the compound;

and determining whether the compound interacts with and activates or inhibits an activity of the polypeptide by detecting the presence or absence of a signal generated from the interaction of the compound with the polypeptide.

18. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with the polypeptide of claim 10, or a fragment or variant thereof, adequate to produce antibody and/or T cell immune response to protect said animal from disease.

19. A method of inducing immunological response in a mammal which comprises delivering a nucleic acid vector to direct expression of a polypeptide of claim 10, or fragment or a variant thereof, for expressing said polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody and/ or T cell immune response to protect said animal from disease.

20. A polynucleotide comprising a polynucleotide sequence selected from the group consisting of the the first ten polynucleotides sequences from the top of Table 1.

21. A polypeptide comprising a polypeptide encoded by the polynucleotide of claim 20.

22. The isolated polynucleotide of claim 1 wherein said nucleotide is selected from the group consisting of:

(a) a polynucleotide having at least a 90% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 90% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 90% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

23. The isolated polynucleotide of claim 1 selected from the group consisting of:

(a) a polynucleotide having at least a 95% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1;

(b) a polynucleotide having at least a 95% identity to a polynucleotide encoding the same mature polypeptide expressed by the gene contained in the *S. pneumoniae* of the deposited strain that was sequenced to obtain a polynucleotide sequence of Table 1;

(c) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 95% identical to the amino acid sequence of Table 1;

(d) a polynucleotide which is complementary to the polynucleotide of (a), (b) or (c); and

(e) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a), (b), (c) or (d).

24. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 50% identity to a polynucleotide encoding a polypeptide comprising the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*;

(b) a polynucleotide encoding a polypeptide comprising an amino acid sequence which is at least 50% identical to the amino acid sequence of Table 1 and obtained from a prokaryotic species other than *S. pneumoniae*; and

(c) a polynucleotide which is complementary to the polynucleotide of (a) or (b).

25. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

26. An isolated nucleic acid encoding one of the amino acid sequences of Claim 1 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

27. Recombinant vectors comprising the nucleic acid sequences of Claim 26 and host cells transformed or transfected therewith.

28. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 1 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

29. Antimicrobial compounds identified by the method of Claim 28.

30. An isolated Streptococcal polypeptide having one of the amino acid sequences given in Table 1.

31. An isolated nucleic acid encoding one of the amino acid sequences of Claim 30 and nucleic acid sequences capable of hybridizing therewith under stringent conditions.

32. Recombinant vectors comprising the nucleic acid sequences of Claim 31 and host cells transformed or transfected therewith.

33. A method of identifying an antimicrobial compound comprising contacting candidate compounds with a polypeptide of Claim 30 and selecting those compounds capable of inhibiting the bioactivity of said polypeptide.

34. Antimicrobial compounds identified by the method of Claim 33.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/21976

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 320.1, 4, 252.3; 536/23.1, 23.7; 530/350, 386; 514/1, 12

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Dialog, APS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y,E	US 5,695,937 A (KINZLER et al.) 09 December 1997, see entire document.	1-34
Y,E	US 5,723,320 A (DEHLINGER) 03 March 1998, see entire document.	1-34
Y,P	US 5,604,100 A (PERLIN) 18 February 1997, see entire document.	1-34
Y,P	US 5,652,128 A (JARVIK) 29 July 1997, see entire document.	1-34

Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means		
"P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
12 MARCH 1998	02 APR 1998

Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer Enrique D. Longton Telephone No. (703) 308-0196	
---	--	---

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/21976

A. CLASSIFICATION OF SUBJECT MATTER:**IPC (6):**

C07H 21/02, 21/04; C12N 15/00; C12P 21/00; C07K 14/00; A61K 35/14, 38/00; C12Q 1/00

A. CLASSIFICATION OF SUBJECT MATTER:**US CL :**

435/69.1, 320.1, 4, 252.3; 536/23.1, 23.7; 530/350, 386; 514/1, 12